THE UNIVERSITY *of* EDINBURGH

Strategies to Identify Novel Therapeutic Targets for Oesophageal Adenocarcinoma

J Robert O'Neill MB. ChB (Hons.), BSc. (Hons), MRCSEd

Thesis Submitted for the Degree of Doctor of Philosophy

The University of Edinburgh

2014

Declaration

I declare that this thesis is the result of my own work and includes nothing which is the outcome of collaboration or external support except where specifically acknowledged in the text.

This thesis has not been submitted as part of any other degree or qualification.

……..

J Robert O'Neill

Abstract

Oesophageal adenocarcinoma (OAC) is a leading cause of cancer death in the UK and current systemic therapies are ineffective for the majority of patients. The central aim of this work was to explore strategies to identify novel therapeutic targets.

Research has failed, thus far, to identify a dominant oncogene in OAC, although the tumour suppressor p53 is frequently mutated. Inhibiting the mitotic kinase, polo-like kinase 1 (PLK-1), was proposed as a synthetic lethal strategy. PLK-1 was demonstrated to be overexpressed in both verified OAC cell lines and human OAC tissue compared to non-transformed cells and epithelium. Mutation of p53 was associated with over-expression of PLK-1 in both OAC and ovarian cancer tissue. Using a carefully validated viability assay, both an established and novel PLK-1 inhibitor were demonstrated to induce a G2/M arrest and reduce OAC cell proliferation. Relative selectivity was demonstrated for OAC compared to non-transformed cells. This therapeutic window could be enhanced with the induction of cancer cell cytotoxicity by pulsed administration of a short half-life inhibitor.

Immunotherapeutics offer potential tumour-selectivity but no OAC-specific proteins have been defined. A comparative proteomic approach was employed to identify OAC-specific proteins as potential therapeutic targets. A tissue resource was established and methods to lyse fresh frozen biopsies optimised. An isobaric quantitative proteomic workflow was applied to OAC and matched normal biopsies and quantitative accuracy confirmed for 6 candidate proteins by immunohistochemistry. Proteome coverage and quantitative dynamic range were compared between isobaric and label-free systematic sequencing proteomic strategies applied to further patients' tissues. The challenges of combining incomplete datasets were approached with a Bayesian framework to estimate the probability that a protein was missed during an experiment compared to not being present in the sample. This method was applied to generate a complete set of protein identifications and relative tissue expression.

To gain insight into the dysregulated cellular processes in human OAC tissue, a network analysis was applied to the quantitative proteomic data. Enriched functional clusters were identified suggesting deranged glucose metabolism, potentially due to the Warburg effect. These findings were duplicated and candidate tumour-specific proteins identified in a further set of biopsies using the optimised quantitative proteomic method. The combined quantitative oesophageal proteomic dataset represents the largest in OAC to date.

This thesis demonstrates a hypothesis-driven, synthetic lethal approach can yield cancer-selective therapeutic effects. Novel candidate therapeutic targets are also revealed through the development of quantitative proteomic methods and the application of network analysis.

Dedication

This thesis is dedicated to my wife, Lynsey.

For tireless support and encouragement.

Acknowledgements

I wish to thank the members of the Hupp lab past and present for bringing me into the fold and for helpful advice, discussion and supervision throughout the project, especially Euan, Erin, Jean, Jenny, Nicky and Terry. Thanks to Tammy Piper for training and advice on histopathology, Dr. Vicki Save and Dr. Rudolf Nenutil for further support with histopathology and slide review.

I wish to thank Dr. Young-Ah Goo and Hui-Song Pak along with Dr. Dave Goodlett, Dr Alex Scherl and Dr. Pavel Bouchal for their help and advice with the proteomic work and for performing the mass spectrometry for this study.

Thanks also to Dr. Ian Overton and Dr. Erola Pairo-Castineira for opening my eyes to the word of computational biology and their patience and support as I took my first painful steps into coding with R and Perl. Thanks to Ian for the suggestion and development of the Bayesian framework for prediction of missing protein identifications.

Thanks to the Cambridge team including Professor Rebecca Fitzgerald for insightful comments and support during this work and Dr. Johnny Ong for help with the immunohistochemistry on tissue microarrays.

The PLK-1 inhibitor used in this project was provided by Dr. Daniella Zheleva (Cyclacel pharmaceuticals).

Finally huge thanks to my supervisor and mentor Professor Ted Hupp. He has been an endless source of ideas and enthusiasm without which none of this would have been possible.

This project was supported by a Wellcome Clinical Training Fellowship through the Edinburgh Clinical Academic Training programme.

Contents

Abbreviations

A_{260}	Absorbance at 260nm
Abl	c-abl oncogene 1, non-receptor tyrosine kinase
APS	Ammonium Persulphate
ACC	Adenocarcinoma
ADCC	Antibody-Dependent Cellular Cytotoxicity
ANOVA	Analysis of Variance
APC	Anaphase Promoting Complex
ATM	Mutated in Ataxia Telangiectasia
ATP	Adenosine Triphosphate
ATR	Ataxia Telangiectasia and Rad3-related protein
AURKA	Aurora Kinase A
BCR-ABL	Breakpoint Cluster Region-V-abl Abelson murine leukemia viral oncogene homolog 1 fusion gene
BAX	BCL2-associated X protein
BIND	Biomolecular Interaction Network Database
BioGRID	Biological General Repository for Interaction Datasets
BO	Barrett's Oesophagus
BPE	Bovine Pituitary Extract
BRAF	v-Raf murine sarcoma viral oncogene homolog B1
BSA	Bovine Serum Albumin
BTK	Bruton agammaglobulinemia tyrosine kinase
CD20	Membrane-spanning 4-domains, subfamily A, member 1
CDK1	Cyclin-dependent Kinase 1
Chk1	Checkpoint Kinase 1
Chk2	Checkpoint Kinase 2
CI	Confidence Interval
CID	Collision-Induced Dissociation
CSF1R	CSF1R – colony stimulating factor 1 receptor
CSNK2A1	Casein kinase 2, alpha 1 polypeptide
Ct	Threshold Cycle
CT	Computed Tomography
DAB	3,3'-Diaminobenzidine
DAVID	Database for the Annotation, Visualisation and Integrated Discovery
DCA	Death Commitment Assay
DDI	Domain-domain Interactions
DFLB	Detergent-Free Lysis Buffer
dH_2O	Distilled Water
DIPS	Database of Interacting Proteins
DLT	Dose Limiting Toxicity

DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DPX	Distyrene Plasticiser and Xylene
DTT	Dithiothretiol
dNTP	Deoxyribosenucleotide Triphosphate
DRAQ5	1, 5-bis,2-di-methylaminoethylamino-4, 8-dihydroxyanthracene-9, 10-dione
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ECL	Enhanced chemiluminescence
ELISA	Enzyme-Linked Immuno Sorbent Assay
EMR	Endoscopic Mucosal Resection
EpiCM-2	Epithelial Cell Media-2
EpiGS-2	Epithelial Growth Supplement-2
ESI	Electrospray ionisation
EUS	Endoscopic Ultrasound
FAM	6-carboxyfluorescein
FASP	Filter-Aided Sample Preparation.
FBS/FCS	Foetal Calf Serum
Fc	Fragment, crystallizable
FDR	False Discovery Rate
FFPE	Formalin-Fixed Paraffin-Embedded
FLN	Functional Linkage Network
FLT	Fms-related tyrosine kinase 1
FLT3	Fms-related tyrosine kinase 3
FLT4	Fms-related tyrosine kinase 4
FOXM1	Forkhead box protein M1
FYN	FYN oncogene related to SRC, FGR, YES
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GOJ	Gastro-oesophageal junction
GORD	Gastro-oesophageal reflux disease
GRP-78	78 kDa glucose-regulated protein
hTERT	Catalytic subunit of the human telomerase gene
HCD	High-Energy Collision Dissociation
HCL	Hydrochloric Acid
HEK 293	Human Embryonic Kidney cell line.
HeLa	Cervical adenocarcinoma cell line.
HEPES-KOH	4-(2-Hydroxyethyl)piperazine-1-ethanesulphonate in potassium hydroxide.
HER2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
HGNC	Human HUGO Gene Nomenclature Committee

hrs	Hours
HRPD	Human Protein Reference Database
HSP-27	Heat shock protein 27
hTERT	Catalytic subunit of the human telomerase gene
hTERT-RPE1	hTERT immortalised retinal pigment epithelium cell line.
HUVECs	Human Umbilical Vein Endothelial Cells
i.v.	Intravenous
IHC	Immunohistochemistry
IntAct	European Bioinformatics Institute Molecular Interaction Database
IPI	International Protein Index
KDR	Kinase insert domain receptor (a type III receptor tyrosine kinase)
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KSFM	Keratinocyte Serum Free Medium
LC	Liquid Chromatography
LR	Likelihood Ratio
MALDI	Matrix-Assisted Laser Desorption Ionisation
MALDI-IMS	Matrix-Assisted Laser Desorption Ionisation - Imaging Mass Spectrometry
MCF10A	Spontaneously immortalised human mammary ductal epithelial cell line.
MCF7	Breast cancer cell line derived from a metastatic pleural effusion.
MCL	Markov Clustering Algorithm
MDM2	Mouse Double Minute 2
MEFs	Mouse Embryonic Fibroblasts
MELK	Maternal embryonic leucine zipper kinase
mFASP	Modified Filter Aided Sample Preparation.
MiaPaCa	Human pancreatic carcinoma cell line
Mins	Minutes
MINT	Molecular Interaction Database
MIPS	Mammalian Protein-Protein Interaction Database
MS	Mass Spectrometry
MS1	Precursor ion mass spectrometry
MS2	Fragment ion mass spectrometry
MS/MS	Tandem Mass Spectrometry
NIH 3T3	Mouse embryonic fibroblast cell line.
nM	Nanomolar
NMWCO	Nominal Molecular Weight Cut-Off
OAC	Oesophageal Adenocarcinoma
OCCAMS	Oesophageal Cancer Clinical and Molecular Stratification
OSCC	Oesophageal Squamous Cell Carcinoma
P21 (WAF1/CIP1)	cyclin-dependent kinase inhibitor 1A,
P53	Protein product of TP53 gene

P/S	0.5 units/ml Penicillin + 500ng/ml Streptomycin
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Ca^{2+} and Mg^{2+} free Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline and 0.1% Tween-20
PCC	Pearson Correlation Coefficient
PCR	Polymerase Chain Reaction
PDGFRB	Platelet-derived growth factor receptor, beta polypeptide
PET	Positron Emission Tomography
PFS	Progression free survival
PI	Propidium Iodide
PLK	Polo-like kinase
PPI	Protein-protein interaction
PTM	Post-translational modification
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNAi	RNA interference
RPMI 1640	Roswell Park Memorial Institute 1640 medium
RPM	Revolutions per minute
RT	Room Temperature
RTK	Receptor Tyrosine Kinase
SAMHD1	SAM domain and HD domain-containing protein 1
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SEM	Standard Error of the Mean
SFPB	Serum Free protein Block
shRNA	Short hairpin RNA
SILAC	Stable Isotope Labeling by Amino Acids in Culture
siRNA	Small interfering RNA
SLB	Sample Loading Buffer
Src	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog
STAT	Signal Transducer and Activator of Transcription
STI	Soyabean Trypsin Inhibitor.
STR	Short Tandem Repeat
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
$t_{1/2}$	Half-life
TBS-T	Tris-Buffered Saline and 0.1% Tween-20
TCEP	Tris-(2-carboxyethyl) phosphine hydrochloride
TEAB	Triethylammonium bicarbonate
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic Acid
TGM-3	Protein-glutamine gamma-glutamyltransferase E

TIFF	Tagged image File Format
TMA	Tissue Microarray
TMT	Tandem Mass Tag
Topors	Topoisomerase I Binding, Arginine/Serine-Rich, E3 Ubiquitin Protein Ligase
TP53	Tumour Protein 53
Tris	Tris(hydroxymethyl)aminomethane
TvO	Ratio of Tumour to Normal Oesophagus expression
TvG	Ratio of Tumour to Normal Gastric expression
µM	Micromolar
UTR	Untranslated region
UV	Ultraviolet
wt	Wildtype
Y2H	Yeast two hybrid assay

Gene name abbreviations used in figures are the Human Genome Gene Nomenclature approved abbreviations. In some pathway diagrams Kyoto Encyclopaedia of Genes and Genomes (KEGG) identifiers have been used. For the sake of brevity the names corresponding to these abbreviations are not listed here but are readily available at http://www.genenames.org/ and http://www.genome.jp/kegg/.

Chapter 1: Introduction

1.1 Oesophageal Cancer

1.1.1 Changes in the incidence of oesophageal cancer and histological subtypes

Oesophageal cancer is the fifth leading cause of cancer death in Scotland and the sixth leading cause worldwide. Globally over 400,000 people die of the disease annually and the incidence continues to increase[1,2]. Oesophageal squamous cell carcinoma (OSCC) is the predominant histological subtype worldwide although there are marked geographical variations in the pattern of disease. Between 1970 and 2000 there was a six-fold increase in the incidence of oesophageal adenocarcinoma (OAC) in the USA and this is now the most common subtype[3]. This marked increase has been mirrored in the UK[4], Europe[5,6] and Australia[7]. Other histological subtypes of oesophageal cancer include neuroendocrine carcinoma, melanoma, lymphoma and leiomyosarcoma but together these contribute less than 5% of all cancer diagnoses. Adenocarcinoma is now the predominant subtype in the UK and this disease will form the focus for this thesis.

Central and east Asia, by contrast, have a higher incidence of OSCC. Discrete geographical areas including northern Iran, northern and central China, southern and eastern Africa have very high incidences of OSCC and this may reflect a combination of genetic predisposition and environmental factors[8].

1.1.2 Effect of Age and Gender on Risk of Oesophageal Cancer

In all countries there is an increasing incidence with age and over 95% of diagnoses are made in people over 45 (Figure 1-1)[1,2]. In Scotland in 2010, oesophageal cancer was diagnosed in 554 men (62%) and 336 women (38%) with deaths from oesophageal cancer in 533 men and 278 women respectively[1]. This male predominance is observed across the UK and across age groups (Figure 1-1). The higher incidence of oesophageal cancer in men is largely due to a marked male predominance for OAC with a male to female incidence ratio of between 4:1 and 5:1[1,9]. The factors responsible for this have not yet been fully elucidated. A greater

prevalence of known risk factors does not appear to completely explain the risk and a protective role for oestrogen and breast-feeding has been proposed[10,11].

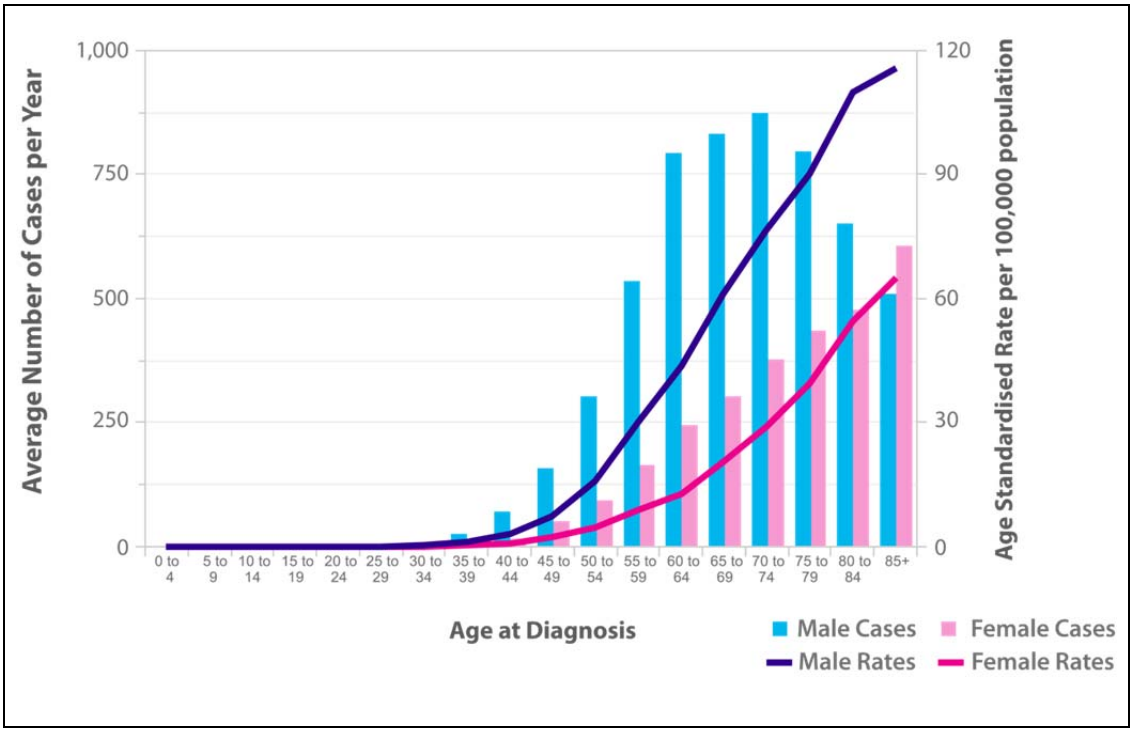

Figure 1-1 The Incidence of Oesophageal Cancer in the UK by Age and Gender (2010 data). Figure adapted from reference[12].

1.1.3 Aetiology and Pathogenesis of OSCC

The oesophagus is lined with squamous epithelium which is directly vulnerable to toxins and viral infection. The role of direct injury in the process of OSCC development is supported by the higher incidence of OSCC in the proximal oesophagus compared to the distal oesophagus. Indeed proximal OSCC shares many risk factors with lung carcinoma including cigarette smoking and excessive alcohol consumption[13,14].

The oesophageal squamous epithelium is also exposed to viruses implicated in oncogenesis. The human papilloma virus (HPV) has been widely implicated in cervical squamous carcinogenesis and evidence of HPV serotype 16 infection has been demonstrated in up to 50% of OSCCs[15]; although there is significant geographic variability in the reported prevalence[16,17].

Further evidence of direct damage in the pathogenesis of OSCC is provided by a case control study from Iran. In the province studied, followers of the custom of drinking scalding hot tea had a significantly higher risk of developing OSCC. This was presumably due to the chronic effects of the direct thermal injury on the upper oesophageal epithelium[8].

The role of alcohol on the development of OSCC appears to be related to the carcinogenic effect of the ethanol metabolite, acetaldehyde[18]. Polymorphisms in the enzymes responsible for ethanol metabolism including ALDH1B and ALDH2 are associated with a four-fold increased risk of OSCC in alcohol drinkers compared to non-drinkers and may explain some of the geographic variation in OSCC incidence[19,20].

Other rarer genetic conditions associated with a high lifetime risk of OSCC include the tylosis oesophageal cancer syndrome which is due to a missense mutation in the RHBDF2 gene leading to altered squamous maturation and potentially altered epidermal growth factor signalling[21,22].

1.1.4 Aetiology and Pathogenesis of OAC

In contrast to OSCC, adenocarcinoma of the oesophagus is most common in the distal oesophagus and at the gastro-oesophageal junction (GOJ). This is likely to be a consequence of one of the major risk factors for OAC; gastro-oesophageal reflux disease (GORD)[23,24]. In one population-based study, patients with recurrent reflux symptoms had a 5-10% increased risk of developing OAC, rising to a >40% increased risk if symptoms were severe and prolonged[23]. Unfortunately, in over 40% of people, significant gastro-oesophageal reflux is asymptomatic[23]. Reflux symptoms are a relatively poor arbiter of risk, however, as most people experiencing reflux do not go on to develop OAC and severe reflux can be asymptomatic in over 40% of people.

Obesity is also associated with an increased risk of OAC[25]. This reflects both the confounding tendency of central obesity to promote GORD[26] but also a potentially pro-inflammatory and carcinogenic effect of visceral adipose tissue[27,28]. This effect

may in part contribute to the gender bias in OAC incidence as men are more likely to accumulate visceral fat than women[29]. Further contributors in obese patients include a low dietary intake of fruit and other anti-oxidants and higher intake of saturated fat, both of which are associated with OAC[24,30].

There is a growing body of evidence suggesting that the combination of refluxed gastric acid and bile creates a distal oesophageal luminal environment most likely to result in the accumulation of genetic mutations and eventually epithelial transformation[31-35]. This same noxious stimulus can, however, lead to epithelial damage and effective mucosal repair, chronic oesophagitis or even loss of the normal stratified squamous epithelium and replacement with a columnar epithelium[36]. The factors governing the oesophageal mucosal response are still poorly understood and likely reflect a combination of host genetic and environmental factors.

Interestingly the development of oesophagitis in response to reflux has not been associated with an increased risk of OAC[37]. It is the appearance of metaplastic columnar epithelium in the distal oesophagus which appears to be associated with the development of OAC[38]. This change was first described over a century ago by the pathologist Wilder Tileston[39] but it was a report in 1957 by the Australian thoracic surgeon, Norman Barrett, that gained recognition and led to the eponym for columnar metaplasia of the oesophagus; "Barrett's oesophagus"[40].

1.1.5 Contemporary Definition of Barrett's Oesophagus.

Three types of columnar epithelium have been described lining the distal oesophagus; a junctional (gastric cardia) type comprised of mucus secreting cells, an acid-secreting gastric type containing parietal and chief cells and intestinal-type metaplasia with prominent goblet cells[41]. These cell types can exist as a mosaic, further complicating diagnosis. In several reports the increased risk of OAC development has been exclusively associated with intestinal-type metaplasia[42,43]. One clinical utility for diagnosing Barrett's oesophagus is to define a group of patients with a higher risk of adenocarcinoma development. Therefore it has been proposed that Barrett's oesophagus should only be diagnosed in patients with intestinal type-metaplasia[44].

This definition has been challenged recently by reports of adenocarcinoma associated with gastric cardia type metaplasia and DNA aneuploidy similar to intestinal metaplasia in cardia-type epithelium[45,46]. Sampling bias can also lead to the under-diagnosis of intestinal-type metaplasia in patients with the characteristic endoscopic salmon-pink mucosal changes of columnar metaplasia[47].

With this background, the British Society for Gastroenterology now defines Barrett's oesophagus as histologically-confirmed replacement of the normal stratified squamous epithelium of the distal oesophagus with metaplastic columnar epithelium of any type, visible endoscopically ≥1 cm above the GOJ[48]. This remains a controversial area and in contrast the most recent American Gastroenterology Association guidelines suggest the presence of goblet cells and therefore intestinal metaplasia are required for a diagnosis of Barrett's oesophagus[49].

1.1.6 Risk of OAC development in Barrett's Oesophagus

Longitudinal studies with serial endoscopic biopsies have identified a pathological sequence of events that are widely regarded to precede the development of OAC. The normal stratified squamous epithelium of the oesophagus undergoes Barrett's metaplasia as previously described. Over time and in the face of further carcinogenic stimuli, foci of dysplasia arise in this metaplastic epithelium. These progress through pathologically distinct low and high grade stages before adenocarcinoma, with the potential for invasion and metastasis, finally develops.

Although Barrett's oesophagus can be considered a premalignant stage of OAC, most patients with Barrett's will not develop adenocarcinoma. Predicting those patients who will progress through the metaplasia-dysplasia-adenocarcinoma sequence has been the subject of intense study for over three decades.

Historical cohort studies have reported the risk of developing OAC with non-dysplastic Barrett's as 30-100 fold increased[50]. However, as endoscopic techniques have improved, the sensitivity to detect dysplasia has increased and the risk for OAC development with a truly non-dysplastic Barrett's segment appears lower[51]. In a Northern Irish cohort with over 8,500 cases, the annual risk of development of high-

grade dysplasia or OAC was 0.38% per year for patients with intestinal metaplasia at index endoscopy[43]. A recent meta-analysis of over 50 studies included 11,434 patients with 58,547 years of follow-up and calculated the risk as 0.33% per year[52].

Although this risk is relatively low, the potential to identify OAC at an earlier, potentially curable, stage has driven the development of Barrett's surveillance programmes. In these, patients with Barrett's oesophagus undergo an upper gastrointestinal endoscopy with targeted biopsy of any mucosal lesions and random mucosal sampling by oesophageal biopsy every three years[48]. The end point of surveillance is normally the identification of high grade dysplasia or adenocarcinoma at which point treatment is instigated.

The risk of progression to adenocarcinoma is significantly higher once dysplasia has developed[43,51]. Unfortunately the classification of low (LGD) and high grade dysplasia (HGD) is based on morphological criteria and is highly subjective. Studies comparing expert upper GI pathologists have determined the kappa value for inter-observer agreement for a diagnosis of LGD can be as low as 0.18[53]. High grade dysplasia (HGD) is associated with a risk of progression to OAC of 20-60% per year[54-56]. Diagnostic concordance is greater for HGD, however, in some series of patients undergoing oesophagectomy for high grade dysplasia, occult adenocarcinoma was identified in up to 40% of cases limiting the application of HGD as a progression biomarker[57,58]. Reproducible biomarkers, evident at earlier pathological stages are clearly required.

1.1.7 Molecular Pathogenesis of Barrett's Carcinogenesis

Oesophageal biopsies obtained from patients undergoing Barrett's surveillance endoscopy and resection for OAC have enabled some of the genetic events present at each stage in the metaplasia-dysplasia-adenocarcinoma sequence to be determined. Although the key features have yet to be determined conclusively, a number features are common to most studies. A selection of these are summarised in Figure 1-2.

Columnar metaplasia is currently believed to arise as an adaptive response to bile and acid reflux in the distal oesophagus. Gene expression studies have identified a key role for the transcription factors Cdx1, Cdx2, BMP-4 and NF-κB in this process[59-61]. This likely reflects the pro-inflammatory effect of bile acid-induced epithelial damage and the subsequent activation of a developmental pathway responsible for intestinal epithelial differentiation[62-64].

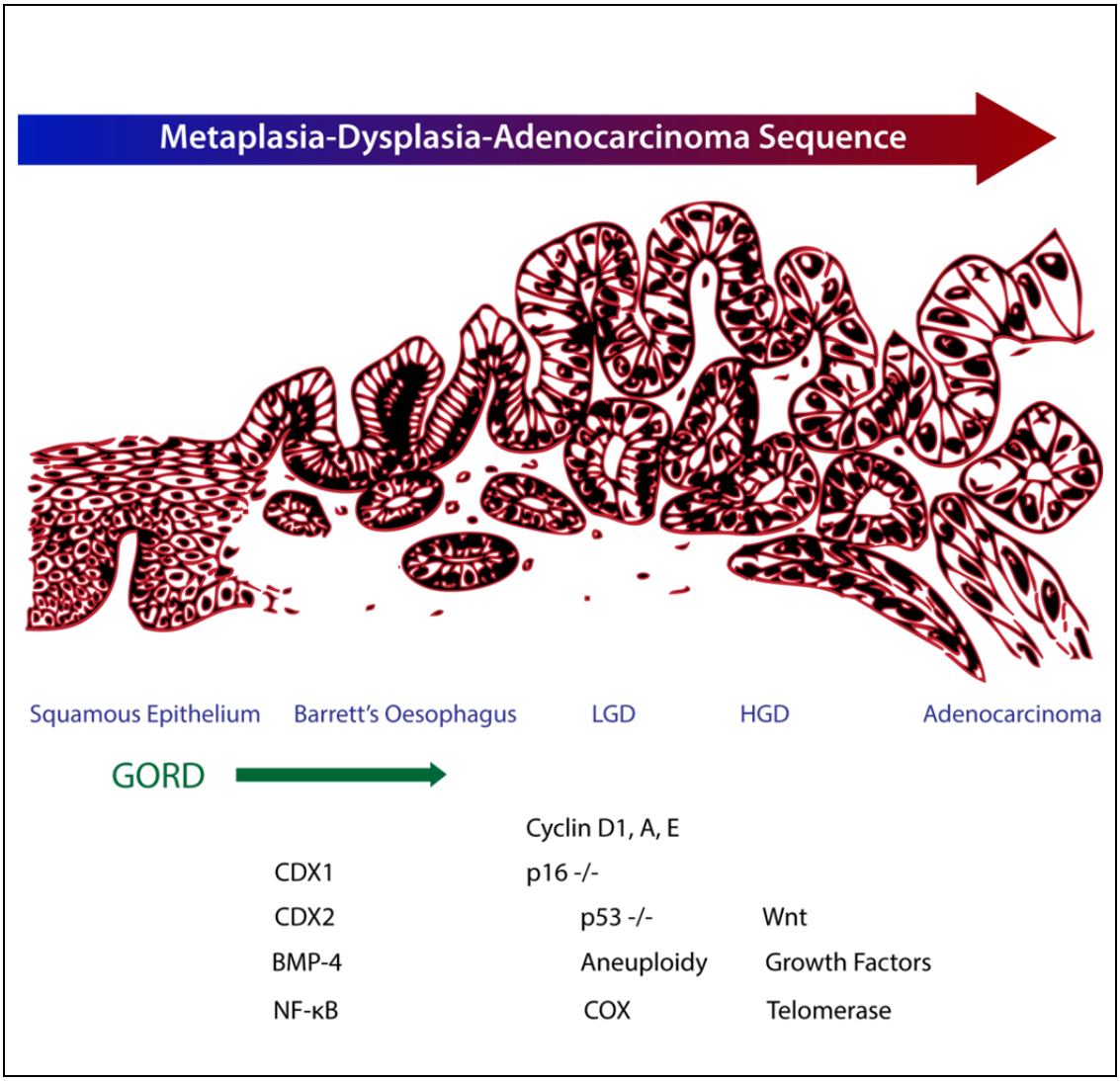

Figure 1-2 Commonly Observed Molecular Changes during Barrett's Carcinogenesis. Adapted from references [65-67]. Gastro-oesophageal reflux disease (GORD) is thought to be a key driver of the adenocarcinoma development. The genes or genetic changes implicated during the progression from Barrett's metaplasia to adenocarcinoma are highlighted at the corresponding stage in the sequence.

During the progression from non-dysplastic Barrett's through LGD and HGD, p16 gene silencing, predominantly by promoter methylation[68-70], and loss of the chromosomal region 17p containing the TP53 gene[71] are key steps. In one series, loss of p16 expression was observed in 85% of patients with non-dysplastic Barrett's[72]. A study reporting the genetic status of individual crypts from Barrett's mucosal biopsies and oesophagectomy specimens confirmed these changes but also unveiled a tapestry of competing clones with heterogeneous p53 mutations within a single Barrett's segment. This highlights the magnitude of the genotoxic oesophageal environment in these patients and the importance of loss of p53 in the progression sequence[73]. The loss of p53 also marks a key step in the development of genomic instability. Aneuploidy together with mutation of p53 have been proposed as biomarkers signifying a higher risk of adenocarcinoma development in patients with non-dysplastic Barrett's[74-76].

In addition to the loss of p53 function, the constitutive activation of Cyclin proteins leads to inhibition of cell-cycle checkpoint mechanisms and clonal proliferation ensues[68,72,77-81]. Ongoing genotoxic stimuli leads to the accumulation of further mutations in these proliferating clones with activation of growth factor pathways including FGF, MAPK, EGFR, TGF-β and Wnt signalling resulting in invasive adenocarcinoma[82-84].

1.1.8 Cell lines as Models of Barrett's and OAC

As the steps during Barrett's carcinogenesis are gradually revealed with more detail, attempts to model the process *in vitro* and in animal systems have intensified. The development of an accurate model would allow the steps of carcinogenesis to be manipulated with potential therapeutic and preventative applications.

After the misidentification of several oesophageal cell lines, a recent study has provided a verified list of OAC lines [85,86]. Ninety percent of these were able to generate tumour xenografts in nude mice and all the xenografts demonstrated the histological features of human adenocarcinoma. These cell lines have been widely

used to assess the effect of chemotherapeutic drugs *in vitro* and to unravel cell signalling mechanisms including the development of treatment resistance[87].

Cancer cell lines are usually grown in monolayers, on plastic plates, with high nutrient growth medium and at a supra-physiological oxygen tension. This has the potential to induce significant changes in phenotype compared to tumour cells in their native microenvironment. This has led some commentators to question the relevance of cell line-based research[88]. Despite these concerns a panel of 51 breast cancer cell lines was found to model the genomic and transcriptomic features of 145 primary breast tumours, including their biological heterogeneity[89]. Importantly, the cell line panel could be used to predict molecular features conferring sensitivity or resistance to a targeted therapy. In a further study using gene expression profiles from 37 gastric cancer cell lines, 2 intrinsic subtypes were predicted. In human gastric tumour samples classified using these molecular subtypes, the groups differed in both prognosis and chemotherapy sensitivity providing direct translational application[90].

As a comparator to cancer cell lines, a series of non-transformed immortalised epithelial cell lines have become available by the stable expression of the catalytic sub-unit of the human telomerase gene (hTERT). These include a Barrett's cell line derived from non-dysplastic intestinal metaplasia[91]. This cell line is already deficient in p16, in keeping with the origin from Barrett's oesophagus. Discrete genetic manipulations, including p53 inhibition and transfection of the oncogenic form of H-Ras, successfully transformed this cell line to allow contact-independent growth, and xenograft development[92]. This highlights some of the key steps required during Barrett's carcinogenesis and this model system may provide a useful tool in therapeutic development.

The parallel use of normal squamous cells allows further comparisons with both transformed cancer cells and Barrett's cells representing an intermediate stage in oesophageal carcinogenesis. Immortalised normal oesophageal squamous cells have been used as a representative model. A widely used squamous cell line, Het-1A, has

been immortalised using the oncogenic virus; SV-40. Unfortunately this immortalises cells by inhibiting p53 and other tumour suppressor proteins[93] and results in a dysplastic growth pattern with a loss of epithelial marker expression when Het-1A cells are cultured in organotypic conditions[94]. For this reason, the Het-1A cell line should not be used as a model of normal squamous oesophagus. A telomerase-immortalised squamous cell line; EPC-hTERT has been described that demonstrates stratified squamous growth in organotypic conditions, or alternatively primary oesophageal squamous cells provide a possible alternative[95].

1.1.9 Animal Models of Barrett's and OAC

Animal models of oesophageal cancer offer the opportunity to closer recapitulate the conditions of human cancer cells *in vivo* and the experimental findings from these models offer potentially greater relevance. One approach is to use nude (immuno-deficient) mice to establish xenografts of human cancer cell lines or fresh explants of human tumour tissue, usually as subcutaneous tumour growths. This approach has the advantage of allowing a tumour mass to grow under similar stresses to human cancer cells *in vivo* but the lack of a competent immune system severely limits the conclusions that can be drawn[96]. This model system is therefore usually reserved to test the effect of a treatment on tumour growth rather than to elaborate the mechanisms of tumour invasion or development.

The majority of animal models trying to recapitulate the development of OAC *de novo* have induced duodeno-gastro-oesophageal reflux through surgical means. A surgical reflux model was first described in dogs but is currently predominantly performed in rodents due to the shorter generation time[97].

The most successful procedure reported is an oesophago-jejunostomy as the presence of both gastric acid and duodenal content in the refluxate is critical for cancer development[98]. Although oesophagitis and columnar metaplasia can usually be generated by this procedure, cancers developing can either be OSCC or OAC. Various dietary supplements and genetic backgrounds have been added to the reflux procedure to address this limitation without consistent success. A further significant limitation of the rodent as a model organism is that the squamocolumnar junction is

in the stomach rather than at the GOJ as in the human[99]. Furthermore both the mouse and rat oesophagus are lined by a keratinized squamous epithelium which does not contain submucosal glands in contrast to the non-keratinized epithelium in humans[100].

This last limitation may be particularly significant as the cell of origin of human Barrett's metaplasia has been proposed as arising in the submucosal ducts of the squamous epithelium[73]. Other possibilities have however, recently been proposed. A study using p63 -/- mouse embryos revealed the development of a Barrett's-like metaplasia in the distal oesophagus[101]. This epithelium demonstrated similar gene expression and immunohistochemical markers as human Barrett's epithelium. Strikingly, lineage tracing suggested this epithelium developed from a small population of embryonic cells present at the squamo-columnar junction. This population was observed in both mouse and human oesophagi. Unfortunately, widespread epithelial defects in the p63 -/- mice preclude their use as models of Barrett's carcinogenesis.

A transgenic mouse model of Barrett's carcinogenesis has, however, recently been proposed. In this model, the Epstein-Barr viral promoter (ED-L2) was used to specifically target the constitutive expression of the inflammatory mediator IL1β to the squamous forestomach, oesophagus and oral cavity[102]. This established chronic oesophagitis which over the course of 20 months progressed through columnar metaplasia to high grade dysplasia and adenocarcinoma in 20% of mice. The process of tumour development could be accelerated by the addition of bile-acids to the drinking water of the mice. Importantly, no squamous cell carcinomas were observed, underscoring the accuracy of the model. As new therapies for OAC develop, this mouse model provides the potential for pre-clinical testing in a representative system.

1.1.10 Diagnosis and Staging of Oesophageal Cancer

Despite increased attempts to identify and survey patients with Barrett's oesophagus, the majority of patients with OAC still present at a symptomatic stage. The oesophagus is a flexible muscular tube and therefore symptoms do not manifest until

the tumour achieves a large enough size to cause obstruction or the tumour invades surrounding structures causing pain, airway compromise or bleeding[103].

Patients are most commonly diagnosed by upper-gastrointestinal endoscopy. This has several advantages over a Barium contrast oesophagogram as biopsies can be taken for confirmation of the histological subtype, the length of the tumour and its relation to the anatomic GOJ can be determined and ionising radiation is avoided[104]. With advances in endoscopy technology, the sensitivity for detection of small cancers has greatly improved and is now superior to contrast radiology.

Controversy exists as to the origin and most appropriate treatment for tumours arising around the GOJ. A subjective clinical classification proposed by Siewert and Stein, has been widely adopted[105]. In this system all tumours centred within 5cm of the GOJ are regarded as junctional. Tumours centred in the distal oesophagus 1-5cm proximal to the GOJ are classified as Type I, Type II tumours are those centred at the anatomical junction, defined as the most proximal extent of the gastric rugal folds and tumours centred 2-5cm distal to the junction are classified as Type III.

Once the diagnosis of oesophageal cancer is made, and if a patient is fit enough to receive further treatment, staging is undertaken. This involves a combination of computed tomography (CT), ^{18}Fluorodeoxyglucose positron-emission tomography (PET), endoscopic ultrasound (EUS) and laparoscopy with peritoneal lavage[106]. The information from these investigations is combined to derive a clinical stage under the International Union against Cancer Tumour Node Metastasis (UICC TNM) classification[107]. This incorporates a prediction of the depth of tumour invasion into the oesophageal wall (T-stage), the extent of regional lymphatic metastasis (N-Stage) and the presence of any distant metastasis (M-Stage) (Figure 1-3). The same classification is used to provide prognostic information on the basis of definitive pathological findings from surgical resections.

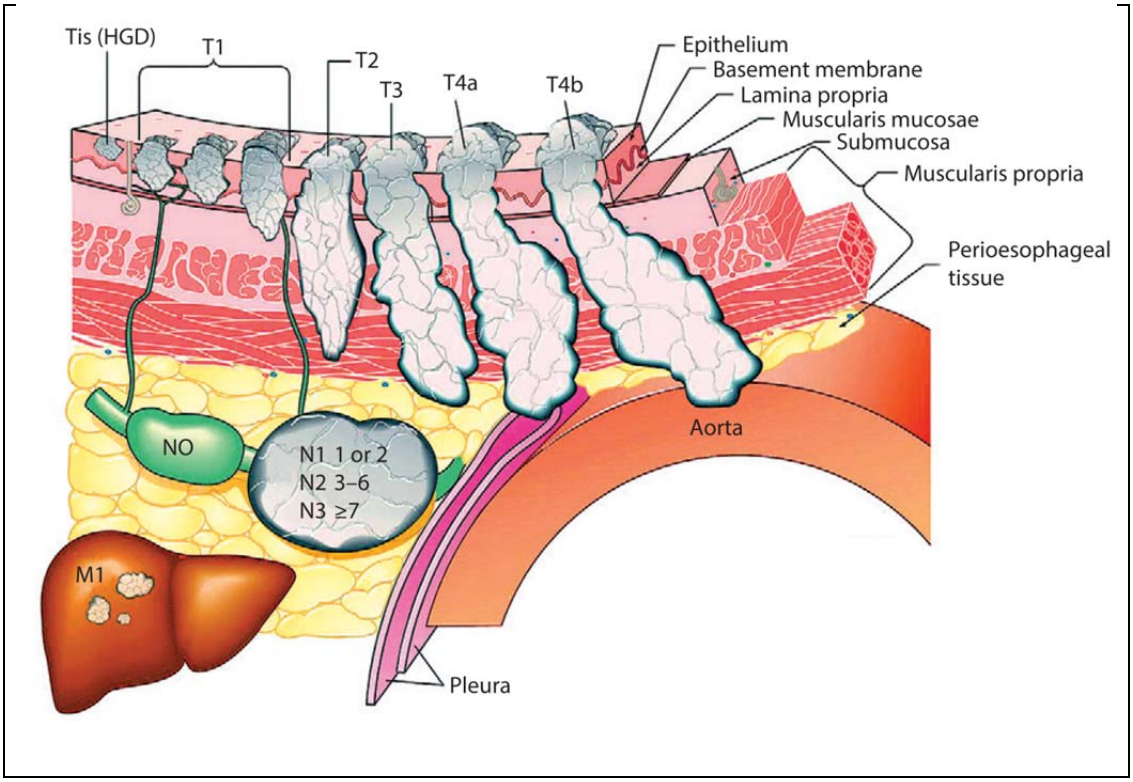

Figure 1-3 Summary of TNM classification for Oesophageal Cancer (7th Edition). From reference[108]

1.1.11 Treatment options for OAC

If staging investigations reveal any evidence of distant metastasis or local invasion rendering the tumour irresectable the patient is considered incurable. Due to the aggressive nature and late presentation of OAC, 70% of patients in the UK have advanced disease at presentation and can only be palliated[1]. For the remainder, attempted curative therapy can be offered and this is tailored to the tumour stage.

1.1.12 Treatment for Early Stage OAC

Patients diagnosed with high-grade dysplasia within Barrett's were historically offered oesophagectomy. This was due to surgical series reporting a combined incidence of occult intramucosal adenocarcinoma of 35% (total n=430)[106]. Although subject to publication bias, only 4 mortalities were reported across these series with excellent oncological results. There is still a significant morbidity associated with oesophagectomy, however, and technological advances including high-resolution white light endoscopy and narrow band imaging have improved the sensitivity for adenocarcinoma detection during endoscopy to greater than 95%[109]. Endoscopic

therapies such as radio-frequency ablation can now be safely applied for HGD, avoiding the need for oesophageal resection although extensive follow-up is still required as the development of metachronous lesions is common[110,111].

Once OAC is diagnosed and metastasis excluded on staging, it is important to determine the T-stage. This is most accurately performed with EUS, although the discrimination between T1 tumours that are entirely intramucosal (T1a) and those that breach the muscularis mucosae to enter the submucosa (T1b) is still poor (see Figure 1-3)[112]. For these early tumours, endoscopic mucosal resection (EMR) is recommended which allows deep biopsies down to the submucosa to permit accurate pathological T-staging[112,113].

The assessment of depth of invasion and delineation between T1a and T1b cancers is critical as the incidence of lymph node metastasis with T1a cancers is less than 2% in published series[114,115]. As tumours invade the submucosa the prevalence of lymphatic metastasis increases from 20% with invasion limited to the superficial submucosal to 50% with deeper invasion[116]. It currently recommended that patients with a T1b or deeper adenocarcinoma are considered for oesophagectomy and lymphadenectomy. In contrast those patients with an intramucosal adenocarcinoma can be potentially cured by EMR if the tumour can be completely excised as the risk of tumour cell dissemination is very low.

1.1.13 Treatment for Locally Advanced OAC

For all attempted curative cancer resections the aim of surgery is to completely remove the primary tumour and the corresponding lymphatic basin[117]. For junctional tumours the lymphatic drainage appears to differ according to the Siewert junctional type. Type II and III tumours predominantly spread to intra-abdominal lymph nodes with only 10% exhibiting lower mediastinal metastasis[118,119]. In contrast, Type I tumours commonly spread to both intra-thoracic and intra-abdominal nodes[120]. The common surgical approaches to these tumours reflect this with a combined thoracic and upper-abdominal lymphadenectomy performed for Type I tumours[121].

Previously, Type I tumours were staged as oesophageal and Type III as gastric with Type II cancers classified according to discretion of the reporting pathologist. The most recent TNM classification has revised this with all junctional tumours involving the oesophagus staged as oesophageal tumours[107]. This classification has been demonstrated to be superior to previous versions in predicting prognosis, most likely due to the revised nodal staging algorithm which incorporates the number of positive lymph nodes[122].

Patients with OAC and regional nodal metastasis undergoing surgical resection alone have a high risk of both local and distant recurrent disease and 5 year survivals remain <20%[122]. In contrast, over 60% of those undergoing surgery alone for OAC with no lymphatic metastasis can expect to survive 5 years[122]. On this basis neoadjuvant therapy is currently recommended for those patients likely to have lymphatic metastasis. From a clinical staging perspective this includes all patients with evidence of lymph node involvement or T3/T4 disease. The management of patients staged as T2N0 remains controversial as clinical staging is inaccurate for three quarters with half of all cases being subsequently upstaged. In many centres these patients are consequently recommended for neoadjuvant therapy[123].

1.1.14 Neoadjuvant therapy for OAC

The MRC OE02 trial randomized 802 patients to receive either neoadjuvant chemotherapy, given as 2 cycles of cisplatin and 5-fluorouracil (2xCF), followed by oesophagectomy or oesophagectomy alone. The 2 year survival rate was 43% in the group receiving chemotherapy compared to 34% in the group undergoing surgery alone (p=0.004). On this basis, neo-adjuvant chemotherapy is recommended in the UK for patients with locally advanced oesophageal cancer undergoing attempted curative therapy. A long-term survival analysis of this trial has recently confirmed the absolute survival advantage of 5% at 5 years although sub-group analysis suggests this is lower in patients with OAC[124].

There are a number of limitations in this trial which reflect the period (1992-1998) during which it was conducted. The post-operative mortality rate was 10% in each group and macroscopically incomplete resections (R2) were performed in 13% of the

surgery only group and 9% of the chemotherapy group[125]. Current data suggest a post-operative mortality rate of <3% should be expected for oesophagectomy in most large-volume centres[126]. Improvements in staging have also reduced the number of R2 resections to <1% of cases[117]. These improvements in surgical care are expected to change outcome and interestingly recent audit data from England and Wales suggests a 2 year survival rate of 56% for patients with OAC and 51% for patients with OSCC, a significant improvement over the OE02 cohort[127].

The OE02 trial findings also contradict the results of the smaller North American Intergroup 113 trial (INT 113)[128]. This trial randomized 440 patients to 3 preoperative and 2 postoperative cycles of cisplatin and 5-FU or surgery alone. This trial had a higher proportion of OSCC (46% compared to 31%) and fewer patients completed the preoperative regime (70% compared to 90% for OE02). Only 40% of eligible patients completed both post-operative cycles perhaps reflecting the challenges of post-operative chemotherapy administration in this patient group. Similar to OE02, 15% of patients underwent an R2 resection in the surgery alone group with 10% in the chemotherapy group. Post-operative mortality rate was 6% in both groups. There was no difference in median survival between groups and 2 year survivals rates were 37% for the surgery only group and 35% for the chemotherapy group. Reasons for this could include the longer delay to surgery for patients not responding to preoperative chemotherapy (12 weeks versus 6 weeks in OE02) and a greater proportion of OSCC masking differences in the OAC groups.

A trial initially powered for patients with gastric cancer (MAGIC) has provided some further evidence of the benefit of pre-operative chemotherapy for patients with GOJ and oesophageal adenocarcinoma[129]. A total of 503 patients were randomized to 3 pre-operative and 3 post-operative cycles of Epirubicin, Cisplatin and 5-FU or surgery alone. Although only 26% of the patients included had GOJ or oesophageal tumours, a trend favouring chemotherapy was noted for both oesophageal and junctional tumours. Pre-operative epirubicin, cisplatin and 5-FU will be compared to cisplatin and 5-FU for patients with OAC or GOJ adenocarcinoma in the OE05 trial[130].

Of note the MAGIC trial again highlighted the challenge of post-operative chemotherapy administration in patients undergoing gastric or oesophageal resection. Although 86% of patients completed pre-operative chemotherapy only 42% completed both all preoperative and post-operative cycles[129]. This reflects the significant morbidity of major upper-gastrointestinal resection and the extended recovery period required.

The use of neoadjuvant radiotherapy alone for oesophageal cancer is not supported by a meta-analysis of randomized trials[131]. Early trials of neoadjuvant chemoradiotherapy were associated with significantly increased perioperative mortality and this approach has not yet been adopted in the UK[132,133]. A recent meta-analysis of trials of neoadjuvant chemoradiotherapy including a total 1,854 patients has, however, identified a 22% reduction in mortality compared to surgery alone (HR 0.78; 95% confidence interval 0.70-0.88, $p<0.0001$)[134]. This strategy has become the predominant treatment for locally advanced oesophageal cancer in Europe, Australia, and North America. A prospective randomised comparison of neoadjuvant chemotherapy and neoadjuvant chemoradiotherapy has yet to be undertaken but would help inform future UK practice[135]. Despite neoadjuvant treatment and surgical resection, over 75% will still succumb to systemic disease and therefore effective therapeutics are urgently required[136].

A complete pathological response (pCR), defined as no viable tumour cells in the resection specimen, has been observed in 25% of neoadjuvant chemoradiotherapy treated patients[137]. Patients experiencing a pCR after chemoradiotherapy have a significant survival benefit over those experiencing no response with 5 year survival rates reported between 34 and 70%[138]. In contrast the pCR rate is <5% across trials of neoadjuvant chemotherapy[139]. Long-term follow-up from the INT 113 trial also suggests a significant disparity in survival between those patients experiencing a response to neoadjuvant chemotherapy and those experiencing no response[140]. Indeed those patients showing no response had a trend for poorer survival than patients undergoing surgery alone. Identifying patients who will go on to respond to neoadjuvant therapy is an active area of research but no biomarkers have yet been validated.

1.1.15 Effect of Immunosuppression during Cancer Therapy

Most chemotherapies in current use act on dividing cells. The toxicity of these drugs can be considerable with treatment related mortality even in the neoadjuvant setting[141]. If these therapies are not effective against the tumour there may be negative consequences to their use.

Surgical resection for OAC carries a risk of major morbidity and the systemic sequelae of surgical stress, particularly catecholamine release may enhance disease dissemination. In a study in rats the use of a β-blocker and a non-steroidal anti-inflammatory drug to mitigate the stress of surgery significantly reduced the development of pulmonary metastasis from intravenous inoculation of a highly metastatic cancer cell line[142]. This effect could be further enhanced by the addition of an immune stimulant; CpG-C. The extension of these observations is that immunosuppression from systemic chemotherapy and surgical stress during tumour resection may allow more effective dissemination of cancer cells.

In OAC, tumour cell invasion and metastasis occur early in the disease process. In a study including 20 patients with the most superficial stage of submucosal oesophageal adenocarcinoma, micrometastasis and isolated tumour cells could be identified by cytokeratin immunohistochemistry in the resected lymph nodes in 20% of cases[143]. Isolated tumour cells could even be identified in the lymph nodes in 8% of patients with intramucosal adenocarcinoma. OAC is also liable to haematogenous spread. Disseminated tumour cells can be demonstrated in iliac crest bone marrow aspirates in a quarter of patients with T1 stage adenocarcinoma undergoing oesophagectomy[144]. The demonstration of disseminated tumour cells is also associated with a significantly poorer survival[144].

The early metastasis of isolated tumour cells in OAC suggests the host immune response to this cancer may be critical in determining if the disease will be indolent or progressive with the development of multiple macroscopic metastases. In breast cancer, circulating tumour cells can also be identified at the time of surgical resection[145]. Although 5 year survival rates of > 90% are expected in stage I and II disease, in up to 20% of these patients, multiple metastases appear over 5 years after

the original treatment[146]. This phenotype has been rationalized with the model of immunosurveillance and tumour cell dormancy.

The immunosurveillance hypothesis states that micrometastases and isolated tumour cells may be maintained in a senescent state by interaction with immune cells. These dormant tumour cells are still viable and can exhibit slow rates of proliferation and cell death. If the balance between immune cell and cancer cell is altered in favour of the latter then rapid proliferation ensues and macroscopic metastases appear[147]. This phenotype has been demonstrated in a mouse breast cancer model[148]. A fatal case of donor transmission of malignant melanoma to a lung transplant recipient despite the donor surviving 32 years after the original melanoma treatment provides additional clinical evidence for this[149]. The powerful immunosuppressive regime required following transplantation presumably allowed tumour cells to rapidly escape their dormant state.

Genetic heterogeneity within solid tumours predicts a small resistant percentage of cells will be clonally selected after most targeted therapies[150]. Tumour selective treatments should however allow tumouricidal doses without the attendant morbidity associated with current therapies. This will also prevent immune compromise, crucially allowing immune-mediated clearance of the residual intrinsically-resistant tumour cells or at least the maintenance of disseminated tumour cells in a dormant state[151]. Identifying tumour-specific therapies and minimising immunosuppression should therefore represent one of the central aims of cancer drug discovery programmes.

1.2 Oncogene Targeted therapy for cancer

The discovery that normal cells could be transformed by oncogenes spawned the targeted-therapy paradigm. This strategy has achieved most success in examples where cells are dependent on the function of one oncogene product. This phenomenon has been coined oncogene-addiction[152]. Notable clinical examples of oncogene addiction have been successfully treated with targeted therapies.

The first successful targeted therapy was imatinib[153], directed against the product of the BCR-ABL fusion gene known to drive the development of chronic myeloid leukaemia[154]. In a landmark randomised controlled trial, Imatinib produced a reduction in BCR-ABL transcript levels of greater than 1000 fold in 39% of cases compared to 2% in the best medical therapy group (interferon + cytarabine)[155]. This resulted in a complete cytogenic remission at 12 months in 68% of patients treated with imatinib compared to 7% of patients treated with interferon + cytarabine.

A subset of non-small cell lung cancers have been found to exhibit oncogene addiction to the epidermal growth factor receptor (EGFR) due to activating mutations within the kinase domain[156]. This has been successfully exploited using EGFR inhibitors such as Gefitinib[157] and Erlotinib[158]. Similarly an activating mutation in another kinase; BRAF has also been identified in over half of all patients with malignant melanoma[159]. This has also been successfully targeted using a specific inhibitor, Vemurafenib which dramatically increased response rates in metastatic disease[160,161].

1.2.1 Targeted therapies for OAC

Following the successes of these targeted agents, researchers have tried to identify similar activated oncogenic kinases in other cancer types. The human epidermal growth factor receptor 2 (HER2) tyrosine kinase is part of the EGF superfamily and is amplified in 20-30% of breast cancers. Trastuzumab is a humanized monoclonal antibody directed against the extracellular domain of HER2 and has been shown to improve disease-free survival in HER2 positive breast cancer[162]. HER2 has also been shown to be amplified in up to 30% of oesophageal and gastric adenocarcinomas[163,164].

A randomized trial of Trastuzumab in combination with chemotherapy versus chemotherapy alone (ToGA) recruited 584 patients with inoperable locally advanced, recurrent or metastatic adenocarcinoma of the stomach (80%) or GOJ (20%)[165]. Trastuzumab significantly improved survival from a median of 11.1 months to 13.8 months (p=0.0046). Consistent with an active effect on HER2, those patients with higher HER2 positivity on IHC demonstrated the greatest improvement in survival. Importantly, there was minimal additional toxicity observed with the addition of Trastuzumab. Following this work, Trastuzumab and a further HER2 targeted agent, Lapatinib, have entered clinical studies in OAC in both the metastatic and perioperative setting and may represent a valuable therapy for HER2 positive tumours[166,167].

Multiple other kinase inhibitors are currently under investigation in clinical trials in OAC either as adjuvants to standard chemotherapy regimens or as monotherapies in patients with metastatic disease. Results, especially for EGFR inhibitors, have been mixed to date[168]. One promising lead is the vascular endothelial growth factor receptor (VEGF) inhibitor bevacizumab. This is currently being evaluated in a phase II/III trial in adenocarcinoma of the stomach and GOJ (National Clinical Trials number 00450203) and also in the neo-adjuvant setting in localised gastric or GOJ adenocarcinoma (National Clinical Trials number 00737438).

Unfortunately gene expression and kinase profiling studies have failed to identify an oncogene or kinase over-expressed in the majority of OACs[60,82,169-171]. Oncogene or kinase-targeted therapies are therefore likely to only be applicable to selected subgroups selected by relevant response biomarkers. An alternative approach that may be relevant to a greater proportion of patients is to target vulnerabilities generated by the loss of a tumour suppressor. This approach, termed synthetic lethality[172], and its relevance to the tumour suppressor p53 will be discussed in the following section.

1.3 Exploiting TP53 Mutation for Cancer Therapy.

1.3.1 The discovery of p53

The human p53 protein was named due to the apparent mass of 53kDa on denaturing polyacrylamide gel electrophoresis (PAGE)[173]. This 393 amino acid protein is encoded by an 11 exon gene located on the short arm of chromosome 17 (17p13.1)[174]. The major functions of p53 appear to be mediated through its role as a transcription factor[175]. The observation of p53 over-expression in transformed cells and binding to viral oncoproteins initially led to its classification as an oncogene[176,177]. A decade after the original identification, the true wild-type TP53 gene was cloned rather than the original sequence which contained a missense mutation. This allowed the true role of p53 as a tumour suppressor to be revealed[178].

1.3.2 The function of p53

The activation of p53 in response to DNA damage leads to expression of genes with important functions in cell cycle arrest, senescence and apoptosis[175]. The relative expression of the genes mediating each function and the subsequent final cellular response is dependent on the mutational and transcriptional state of the cell. The fine details of the regulatory mechanisms are still being uncovered and as the ability to identify transcriptional targets improves, the number of p53-responsive elements, and potential p53-mediated functions, has expanded considerably[179]. Despite this increased complexity, the central role of p53 in the response to DNA damage and designation as “guardian of the genome” underpins the likely reason this gene is frequently mutated in cancer[180,181].

1.3.3 The Domain Structure of p53

Five functional domains within the protein have been described (Figure 1-4). The transactivation domain, immediately adjacent to the proline-rich domain in the amino-terminus of the mature protein contains a sequence that allows efficient activation of transcriptional targets[182]. The proline rich domain is thought to contribute to the p53-mediated apoptotic response and the transcriptional response to gamma irradiation[183]. Post-translational modification of p53 can augment or inhibit

its interactions with down-stream targets and the transactivation and proline-rich domains are key sites for reversible phosphorylation[184].

The DNA-binding domain is the most frequent site of TP53 mutation in cancer. Some mutations have been identified at a far higher frequency than others; so-called "hot-spot" mutations. These have been subsequently demonstrated to coincide with key residues responsible for DNA-binding, consistent with the role of p53 as a transcription factor[185]. For example, Arginine 248 is critical for binding to the minor groove of the DNA double helix and mutation of this residue is the most commonly observed p53 mutation in cancer[186]. The function of p53 has also been demonstrated to be dependent on the assembly of a p53 tetramer via the tetramerisation domain, and this domain is also a site of mutation in cancer[187,188].

The c-terminal regulatory domain is thought to modulate the function of p53 in response to external factors and can switch p53 from a latent to an active DNA-binding conformation. This may occur via an allosteric mechanism where the DNA-binding domain availability is decreased within the tetramer by regulatory domain binding[189,190].

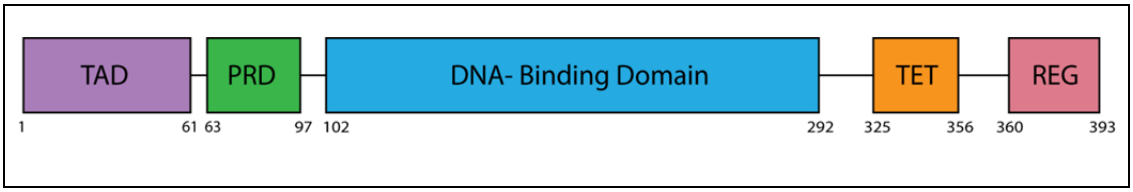

Figure 1-4 The domain structure of human p53 protein. The amino acid positions are detailed below. TAD – Transactivation domain, PRD – Proline rich domain, TET – Tetramerisation domain, REG – regulatory domain.

1.3.4 MDM2 and p53

One of the transcriptional targets of p53 is the E3 ubiquitin ligase MDM2 which inhibits the function of p53. This occurs by both direct binding and inhibition of transactivation, and through enhanced p53 degradation at the proteasome via MDM2-mediated polyubiquitylation[191,192]. This negative feedback loop ensures expression of p53 protein is transient to prevent deleterious activation of apoptotic machinery. Conversely, stable expression of p53 is often observed in cancers due to

the loss of MDM2 transactivation, and subsequent p53 degradation. Overexpression of p53 protein has therefore been used as a surrogate marker of loss of p53 function[193].

1.3.5 TP53 is commonly mutated in cancer

The combined evasion of growth suppression and inhibition of apoptotic signalling confers a significant survival advantage on cells with loss of p53 transcriptional activity[186]. In the process of oncogenesis, clones with mutant p53 may be selected due to evolutionary pressure and therefore p53 is frequently observed as a mutated gene in cancer[186].

Large scale sequencing efforts have now confirmed a high frequency of p53 mutation in several cancer types including head and neck squamous cell carcinomas (SCC), endometrial cancers and high grade serous ovarian cancers[194-197]. In head and neck SCCs, mutations in TP53 were identified in 50 – 80% of the tumours with evidence of human papilloma virus 16 infection in wild-type tumours[196,197]. This confirms the loss of p53 function as critical in SCC development as the E6 viral protein induced by HPV infection has transforming potential through the inactivation of wild-type p53 protein[198].

A further striking example of the selective pressure for mutation of p53 in the development of cancer comes from a laser capture microdissection study on Barrett's epithelium[199]. This revealed multiple clones with differing p16 and p53 mutations arising from the same Barrett's field confirming the mutagenic environment of the distal oesophagus in these patients. In some cases, a single p53 mutation was observed over a wide epithelial area suggesting rapid expansion of this clone due to a competitive advantage.

1.3.6 Mutant p53 gain of function

Some p53 mutations are associated with oncogenic functions including the ability to complex and inactivate other p53-related proteins integral to the tumour suppressor mechanism including p63 and p73[200]. Disruption of the mutant p53-p73 interaction inhibits cell growth *in vitro* and the ability to form tumour xenografts *in vivo*[201].

Mutation of one allele of p53 can also lead to a loss of p53 function due to inactivation of the wild-type protein via a dominant negative effect[202].

Selective knockdown of overexpressed mutant isoforms by RNA interference increases chemo- and radiosensitivity, induces apoptosis and reduces tumour cell growth and capacity to metastasise *in vivo* suggesting mutant p53 may be a potential therapeutic target in cancer[203,204]. Small molecules have been also been developed to target mutant p53 proteins[205] although targeting the thousands of discrete mutant isoforms directly may prove difficult with a single drug[206].

Different p53 mutations may confer different phenotypes including variable chemotherapy sensitivities[207]. Mutant p53 has been associated with poor survival in oesophageal adenocarcinoma[208] and mutations in part of the DNA binding domain of the protein are associated with poor prognosis in colon cancers[209]. Targeted mutagenesis has been used to catalogue the phenotype of the heterogeneous p53 mutations[210].

1.3.7 TP53 is Commonly Mutated in OAC

Multiple studies have assessed the mutation frequency of p53 in oesophageal adenocarcinoma. Historically this was achieved by identifying coding variants within exons 4-9 by sequencing of restriction fragment length polymorphisms[206,211]. Reported frequencies using this method include 15/17 cases (88%)[212], 18/37 cases (49%)[213] , 67/142 cases (47%)[208], 8/9 cases (89%)[214], 17/30 cases (57%)[215] and 30/40 cases (75%)[216] producing a total of 155/254 cases (61%). This method will miss mutations outside the sequenced exons and will also miss a functionally inactivated pathway with a wild-type p53 allele. Notably these studies also report p53 over-expression in supposedly wild-type tumours. This supports the presence of occult mutations or pathway inactivation by some other mechanism[217]. Studies examining p53 protein over-expression (a surrogate marker of pathway inactivation) report frequencies of ~80%[217].

One recent study has sequenced the exomes of 149 OACs and patient-matched normal tissues at over 80-fold coverage, providing a very high sensitivity for

identification of TP53 genetic variants[218]. Somatic p53 mutations were detected in 108 tumours (72%) confirming the high frequency reported in previous smaller studies.

1.3.8 Loss of tumour suppressors provides a synthetic lethal opportunity

Loss of tumour suppressor activity can generate unique dependencies on targetable pathways. The synthetic lethal hypothesis states that inhibiting these pathways can yield cancer-specific cytotoxic effects. This approach is effective in cancers deficient in Breast Cancer proteins (BRCA)1 and 2 where sensitivity to poly(ADP-ribose) polymerase-1 (PARP-1) inhibitors has been described[219]. A similar approach may be feasible in tumours with mutant p53.

A study on the effect of p53 loss on gene expression used isogenic HCT116 cell lines with either wild type p53, heterozygous (p53 +/-) or homozygous (p53 -/-) loss of expression of the endogenous protein[220]. Using gene expression profiling, the HCT116 p53 -/- cell line was found to express higher levels of genes associated with the G2/M checkpoint, in particular Polokinase-1 (PLK-1)[220]. This prompts the suggestion that loss of p53 may be associated with unique dependency on genes involved in this checkpoint.

This hypothesis is supported by a large-scale transcriptomic and genomic study of high-grade serous ovarian adenocarcinomas[195]. This cancer was characterised by a TP53 mutation frequency of 96% suggesting this is a critical step in the development of this cancer. Using gene expression data, copy number analysis and a computational pathway activity inference method[221] a significantly dysregulated network of genes centred on the transcription factor FOXM1 was identified. This network is responsible for the control of various phases of the cell cycle including G2/M exit and may represent a therapeutic target[222].

These data, both from *in vitro* and *in vivo* studies support a model where loss of p53 function creates upregulation of genes related to the G2/M checkpoint. Targeting the cell cycle may therefore represent a therapeutic strategy in tumours with mutant p53.

1.3.9 The Cell Cycle

All dividing cells possess sophisticated cell cycle machinery to allow the perfect replication and propagation of the genome. A carefully orchestrated, kinase-mediated sequence of events is responsible for driving cell cycle progression. The core features of this cycle are conserved across prokaryotes, yeast and diverse animal phyla. An uncorrected error at any stage of this process could allow the transmission of a corrupt genetic code to daughter cells with dire consequences. Checkpoints have evolved within the cell cycle to allow error- correction or, in the case of irreparable damage, for the cell to enter a state of senescence or undergo apoptosis[223]. The task of faultless DNA replication is further complicated by mutagens including UV light, ionising radiation, reactive oxygen species and chemical carcinogens inducing DNA damage.

The cell cycle has been conceptually subdivided into four phases based on two processes, DNA synthesis (S phase) and mitosis (M Phase) separated by two gaps (G1 and G2). The principle checkpoints exist at the G1/S phase transition, intra-S phase, G2/M phase transition and at the spindle assembly stage of mitosis. A major checkpoint trigger is DNA damage and although checkpoint activation is the final endpoint, different forms of DNA damage appear to induce different signalling pathways.

The double-stranded DNA breaks produced by ionising radiation predominantly induce the Mutated in Ataxia Telangiectasia (ATM) pathway and the interstrand cross-links and single-strand DNA breaks produced by UV light (and Cisplatin) predominantly induce the ATM and Rad3 related protein (ATR) pathway[224]. These pathways both comprise of a canonical series of phosphorylation events that finally inactivate the cyclin dependent kinases (CDK) and arrest the cell cycle.

Part of the ATM and ATR-dependent DNA damage response leads to phosphorylation and activation of the tumour suppressor p53 with subsequent expression of the CDK inhibitor p21 and G1-S phase checkpoint activation. The presence of a functional p53 pathway appears to be critical for initiation and maintenance of G1-S phase arrest after DNA damage[220,224]. One of the key roles of

p53 is to trigger the cell cycle checkpoints at the G1/S and G2/M transitions. These checkpoints allow DNA repair and without them cells rapidly accumulate lethal genetic mutations and undergo cell death[225]. With the loss of p53 function, cancer cells become dependent on other mechanisms for checkpoint activation and these mechanisms may represent key targets for synthetic lethality [226]. The G2/M checkpoint can, for example, be activated independent of p53 and may become the critical point of DNA repair in cells without wild-type p53. This is supported by the finding that while p53 is the most commonly mutated gene in human cancer, genes involved in the G2/M checkpoint are rarely mutated[227].

1.3.10 The G2/M checkpoint

The Chk1 and Chk2 kinases are key mediators of the G2/M checkpoint and are activated by ATM/ATR in response to DNA damage[223]. These kinases inactivate cdc25A-C preventing the activation of the cyclin B/CDK1 complex and delaying the onset of mitosis. An essential component in triggering the G2/M checkpoint is the inhibition of the mitotic kinases; Polo-like kinase 1 (PLK-1) and Aurora kinase A (AURKA).

1.4 Polo-like Kinase 1: A Mitotic Kinase

1.4.1 The Polo-like Kinase Family.

A search for mitotic mutants in *Drosophila* larval neuroblasts identified the *polo* gene[228,229]. The human ortholog of *polo*, Pololike kinase 1 (PLK-1), was cloned three years later[230]. A further 4 human polo-like kinase genes have now been described. PLK-1 to 4 share structural homology with an N-terminal serine/threonine kinase domain and one (PLK-4) or two c-terminal polo-box domains (PLK-1 to 3)[231]. While the kinase domain is responsible for the catalytic function of these proteins, it is structurally similar to many other kinases and substrate specificity appears to be controlled by the polo-box domain (PBD)[232]. A fifth PLK family member (PLK-5) has been identified with a homologous PBD but it does not contain a kinase domain and appears to control neuron differentiation rather than function in the cell cycle[233].

1.4.2 A Mitotic Role for PLK-1

PLK-1 functions as a master switch controlling entry to mitosis and recovery from the G2/M checkpoint[234-236]. In unstressed conditions, cells can proceed to mitosis in the absence of PLK-1. After DNA damage, however, active PLK-1 becomes essential for initiation of mitosis and therefore PLK-1 is critical for cell cycle re-start following G2/M arrest [236-238]. PLK-1 appears to be integral in centrosome maturation, spindle assembly, chromosome segregation and cytokinesis[239-244]. The role of PLK-1 in mitosis has recently been comprehensively reviewed[245]. PLK-1 has also been implicated in the AMP-activated protein kinase pathway and preliminary studies propose links to cell motility through changes in γ-tubulin localisation[246].

1.4.3 Regulation of PLK-1 expression

PLK-1 expression is tightly controlled to provide appropriate expression given the cell cycle context. In normal cycling human cells, expression rises through s-phase to peak at the G2/M transition[247]. The major positive regulators of transcription are the forkhead transcription factors - in particular FOXM1[248]. These transcription factors act in a positive feedback loop with both PLK-1 and FOXM1 further enhancing FOXM1 transcription[249,250]. Levels are maintained through mitosis and then, between late M phase and early G1, PLK-1 is ubiquitylated, potentially by the anaphase promoting complex (APC), and degraded at the proteasome[251]. PLK-1 expression is suppressed during G1 phase by the cell-cycle-dependent element/cell cycle gene homology region protein (CDE/CHR) allowing the cycle to reset[247].

1.4.4 Regulation of PLK-1 activity

The kinase activity of PLK-1 is itself regulated by reversible phosphorylation. In the inactive conformation, the C-terminal PBD associates with the kinase domain preventing ATP-binding[231]. Phosphorylation of threonine 210 in the T-loop of the PLK-1 kinase domain is associated with a conformational change which allows the polobox and kinase domains to dissociate and dock with their respective ligands [252]. The activation of PLK-1 in this fashion is controlled by Aurora kinase A and the co-factor hBora[236,238]. Activation of PLK-1 allows substrate recognition and reversible phosphorylation of the key PLK-1 cell cycle targets Wee-1 and cdc25A[236]. The

combined inactivation and activation of these targets drives dephosphorylation and activation of the Cyclin B/CDK1 complex, restarting the cell cycle machinery. The reciprocal dephosphorylation and deactivation of PLK-1 is controlled by protein phosphatase 1 (PP1) and the co-factor myosin phosphatase-targeting subunit 1 (MYPT1)[253] (summarised in Figure 1-5).

1.4.5 Checkpoint Adaptation.

Cells sustaining DNA damage during G2 can trigger the G2/M checkpoint and engage DNA repair mechanisms. If the damage is corrected then cells can proceed onto mitosis. If the damage cannot be corrected the cells can trigger apoptosis or remain at the checkpoint. Adaptation to the checkpoint has been described *in vitro*. In this process, cells enter mitosis after a prolonged G2 arrest such as that induced by irradiation and demonstrate persistent DNA double-stranded breaks. This adaptation can be enhanced by depletion of Chk1 and inhibited by depletion of PLK-1[254]. It is possible that dysregulation of this pathway due to hyperactive PLK-1 could promote early mitotic onset in the presence of damaged DNA leading to genomic instability. Aneuploidy is a general feature of oesophageal adenocarcinomas and is also observed in Barrett's, predicting a high risk of progression to adenocarcinoma[68,255].

Checkpoint adaptation may allow gradual progression of carcinogenesis due to tolerated genomic instability. This may be exploited in cells with mutant p53 where only the G2/M checkpoint is functional. Complete inhibition of this checkpoint sensitises cancer cells to DNA damage and can induce massive apoptosis when combined with a genotoxic treatment. This has been demonstrated with selective Chk1 inhibitors in a cervical, lung and colon cancer cell lines[256]. A role for a Wee-1 inhibitor as a radiosensitising agent has also been proposed in glioblastoma[226].

In glioblastoma cells with wt p53, induction of DNA damage with Temozolomide induced a G2/M arrest which was preserved for over 7 days. Subsequent to this, cells became senescent. When p53 function was inhibited in the same cells, the G2/M arrest was followed by apoptosis and a loss of colony formation capacity[257]. This provides evidence that the fate of cells after prolonged G2/M arrest may depend on p53 function. Inhibiting the drivers of G2/M exit, the mitotic kinases, may

therefore allow a therapeutic window to be achieved between cells differing in their p53 status. The key mitotic kinase responsible for the G2/M exit is PLK-1.

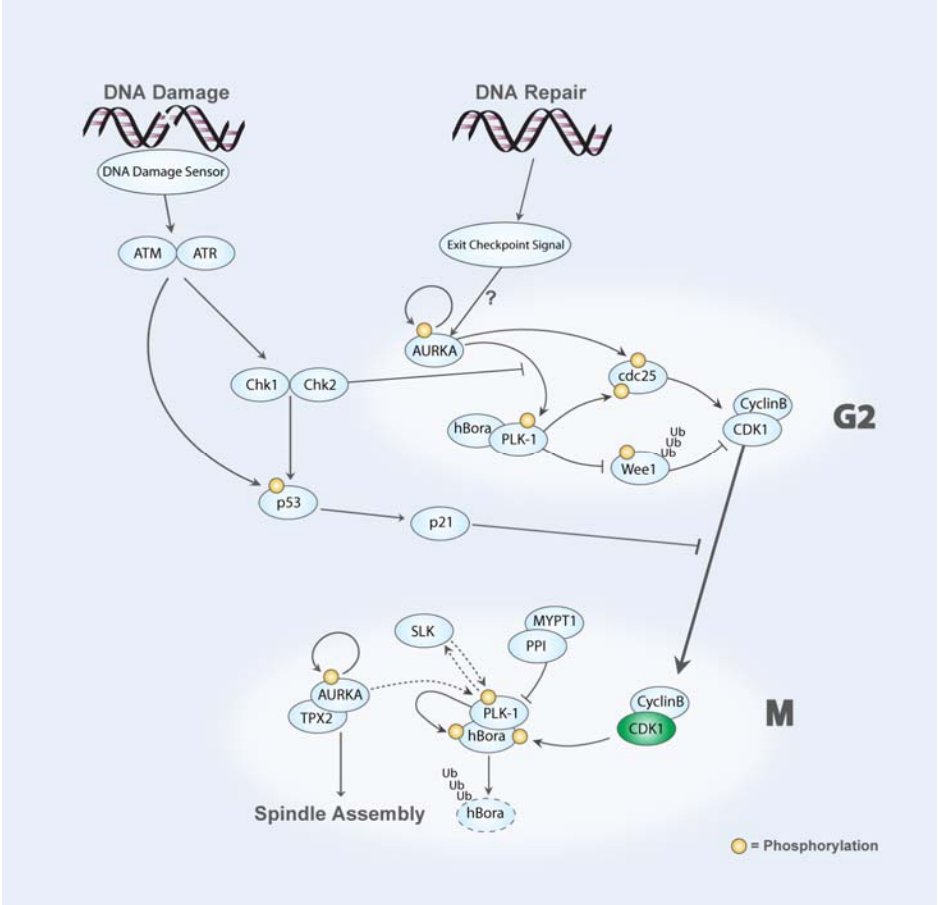

Figure 1-5 Key Components in the G2/M Transition. Adapted from Reference[237]. Damaged DNA is recognised by the DNA damage sensor mechanism consisting of a complex of proteins, specific to the type of DNA damage including double and single stranded DNA breaks and replication stalls. A phosphorylation cascade is triggered inducing the activation of ATM/ATR and Chk1 and Chk2. This leads to stabilisation and activation of p53 via phosphorylation. In cells with functional p53, p21 is upregulated and leads to G2 arrest by direct inhibition of the Cyclin B/CDK1 complex[258]. A p53-independent checkpoint mechanism can also be triggered by the Chk1/Chk2-mediated inhibition of PLK-1 activation. This prevents the phosphorylation-dependent recognition of Wee-1 by the E3 ubiquitin ligase SCF-βTRCP1/2 and subsequent proteasomal degradation. The abundant Wee-1 can then phosphorylate and inactivate the CDK1-CyclinB complex leading to G2 arrest[259].

After DNA repair, a still elusive signal triggers checkpoint exit. Aurora A becomes activated and together with the co-factor hBora phosphorylates PLK-1 on threonine 210[238]. Phosphorylated PLK-1 then phosphorylates and activates the phosphatase Cdc25[260] and targets wee1 for ubiquitylation and degradation. The cyclin B/CDK1 complex is activated and triggers mitotic entry. Active PLK-1 creates a phosphodegron on hBora which allows recognition by the SCF- βTRCP1/2 complex, polyubiquitylation and degradation[261]. Aurora kinase A is then released from the hBora-PLK-1 complex and together with the co-factor TPX2 becomes critical for the initiation of bipolar spindles. Ste20-like kinase (SLK) may play a role in the AURKA-independent maintenance of phosphothreonine 210 Plk-1 in mitosis. With mitosis complete, phosphothreonine 210 Plk-1 is degraded by interaction with protein phosphatase 1(PP1)-Myosin phosphatase-targeting subunit 1 (MYPTY1) and the cycle resets.

1.5 The role of Polokinase -1 in Cancer

1.5.1 PLK-1 Expression in Cancer

PLK-1 overexpression has now been demonstrated in a large number of cancer types and it appears PLK-1 mRNA levels are elevated in the vast majority of rapidly dividing cells[262,263]. Expression has been correlated with poor prognosis in breast cancer[264], lung cancer[265], head and neck cancer[266] among others. PLK-1 also appears to be independently oncogenic as over-expression can transform the spontaneously immortalised mouse fibroblast cell line NIH 3T3[267].

1.5.2 PLK-1 Expression in Oesophageal Squamous Cell Carcinoma

PLK-1 mRNA has been shown to be overexpressed in OSCC tissue compared to normal squamous epithelium by several authors and has been independently associated with a poorer prognosis[262,268-270]. When multiple OSCC cell lines and the OAC lines OE33 and FLO-1 were subjected to limited gene expression profiling using a custom cDNA array, PLK-1 was significantly overexpressed compared to a mixture of human normal tissues[271]. This overexpression was confirmed by quantitative real-time polymerase chain reaction (qRT-PCR) in OSCC and OAC tissues compared to a wide range of normal tissues with the exception of testis which expressed higher levels of PLK-1 than either cancer type. The PLK-1 promoter was also confirmed to be more active in OSCC cells compared to normal human bronchial and renal cortical epithelial cells although cell proliferation rate may be the explanation for this[271].

PLK-1 has also been shown to be overexpressed in OSCC compared to adjacent normal tissues by IHC using a tissue microarray with material from 150 patients[269]. These findings were reproduced in a further study of 79 patients[272]. In both studies, PLK-1 expression was restricted to a few cells within the proliferating basal zone in oesophageal squamous epithelium but was more widely expressed in OSCC. High PLK-1 protein expression was also correlated with lymph node metastasis, and independently predictive of survival[269]. Using fluorescence *in situ* hybridisation to assess PLK-1 copy number, the same authors reported amplification in 6 out of 16 cases[269]. This is the only report to suggest PLK-1 amplification as a mechanism of

overexpression in cancer and no data were presented for control probes against other genomic loci. Therefore this can only be interpreted with caution since aneuploidy, a common finding in this cancer, rather than PLK-1 locus amplification, may explain the findings[273].

Rodent reflux models have also been used to generate oesophageal tumours bearing a mixture or adenocarcinoma and squamous cell carcinoma histologies. When three cell lines derived from these reflux-induced tumours were xenografted into nude mice, well differentiated squamous cell carcinomas were produced. PLK-1 was found to be highly expressed in these cell lines further underlining its prominent role in either the development or maintenance of OSCC[274].

1.5.3 PLK-1 Expression in OAC

A study of gene expression in oesophageal and OGJ adenocarcinomas identified PLK-1 mRNA expression as strongly correlated with the number of lymph node metastasis, the most significant determinant of prognosis in these patients[275]. PLK-1 mRNA overexpression has also been demonstrated in OAC tissue compared to Barrett's and normal squamous oesophageal tissue although the results were not confirmed by IHC[276]. Overexpression of PLK-1 in OAC is closely correlated with FOXM1, cyclin B1, Aurora kinase B and ki-67. These genes form part of a transcriptional network closely linked to mitotic progression and their overexpression was associated with a poorer prognosis in OAC[276,277].

1.5.4 Inhibiting PLK-1

The evidence for PLK-1 as a potential driver of cancer cell proliferation prompted research into the effects of PLK-1 inhibition. This was initially accomplished by genetic means, including antisense oligonucleotides, transfection of a dominant negative mutant and RNA interference[260,278-286]. Interest from the pharmaceutical industry in PLK-1 as an anticancer target prompted the development of selective inhibitors of the PLK-1 kinase domain. The high level of sequence and structural similarity between the kinase domains of pololike kinase family members presents a significant challenge for the development of isoform specific inhibitors[287-289]. The first potent PLK-1 inhibitor developed, for example, inhibits pololike kinases with

over 1,000-fold greater potency compared to a panel of over 60 other human kinases, but only inhibited PLK-1with 4 fold greater potency than PLK-2 and 11 fold greater potency than PLK-3[290]. This may have a significant impact on the *in vivo* effect of this compound as PLK-2 and PLK-3 have potentially antagonistic effects to PLK-1.

In contrast to the tight regulation of PLK-1 expression, both PLK-2 and PLK-3 are rapidly and transiently expressed after a diverse range of external stimuli including growth factor or hormone exposure and cellular stress as part of a group of immediate early response genes[291,292]. PLK-2 is predominantly expressed during G1 and exhibits tumour suppressor properties[293]. PLK-3 is activated in response to DNA damage and is reported to activate and stabilise p53 by phosphorylation at serine-20[294]. With this in mind, PLK-1 inhibitors that have been optimised to enhance the specificity for PLK-1 may possess greater *in vivo* activity (Table 1-1).

1.5.5 Enhancing Specificity for PLK-1

Several approaches have been attempted to enhance drug specificity for PLK-1. Non-ATP competitive inhibitors thought to inhibit PLK-1 kinase activity include ON01910a[295] and Cyclapolin 1[296]. ON01910a is reported to produce a G2/M accumulation and multipolar spindles and appears to be relatively selective *in vitro* for PLK-1 over other pololike kinases[295]. This inhibitor is, however, active against a range of other kinases and the mitotic phenotype differed from cells treated with PLK-1 RNAi prompting the suggestion this inhibitor does not act via PLK-1 inhibition[297]. Cyclapolin 1 was identified through structure-guided drug design and inhibits the kinase activity of PLK-1 in a non-ATP competitive mechanism which may involve covalent modification of a residue in the ATP binding cleft[296]. It is unknown if this mechanism of inhibition confers greater specificity as the activity against PLK2 and PLK3 has not been reported for this compound.

A novel strategy to inhibit PLK function *in vivo* is to target the ligand-binding interface of the PBD to prevent localisation of the kinase domain with biological substrates. This structure is unique to pololike kinases and therefore off-target effects across the kinome can potentially be avoided. The first in class non-peptide PBD inhibitor, poloxin, exhibited specific activity against pololike kinases[298]. The

phosphopeptide binding groove of the PBD differs between PLK-1,2 and 3 and therefore the potential for isoform specific inhibitors appears greater. This inhibitor, however, only showed a 4 fold relative specificity for PLK-1 over PLK-2. Treatment with this inhibitor induced mislocalisation of PLK-1, mitotic arrest, and apoptosis in Hela cells consistent with a specific effect on PLK-1. Similar effects were also observed with the naturally occurring compound Purpurogallin, and the non-selective PBD inhibitor Poloxipan[299,300]. These phenotypes were reproduced in a further study through overexpression of the PBD domain or expression of a kinase-domain only PLK-1 mutant[301]. This confirms the subcellular localisation of PLK-1, mediated via PBD interactions, is critical for PLK-1 function.

1.5.6 Early Clinical Trials of PLK-1 Inhibitors

Encouraging preclinical data has propelled several PLK-1 inhibitors into clinical trials (Table 1-2). The observed major toxicities appear to be related to the activity of the inhibitors on haematopoietic cells with dose limiting neutropaenia and leucopaenia in the majority of phase I trials. This haematological toxicity can be a significant source of morbidity and even mortality but can be controlled in the clinical setting of advanced malignancy if therapeutic efficacy is observed. Unfortunately the strong preclinical anti-tumour activity of the PLK-1 inhibitors has not been replicated in Phase II studies of solid cancers to date. Partial response rates, commonly determined by cross-sectional imaging of tumour deposit size, of <5% of treated cases have been reported with BI2536 (Table 1-2). The relatively limited anticancer efficacy in the early clinical studies of PLK-1 inhibitors highlights the need to identify predictors of sensitivity and methods to enhance the therapeutic index between cancer and normal tissues.

Table 1-1 Published PLK-1 Inhibitors. Adapted in part from reference[302].

Compound	Mechanism of Action	IC_{50} for PLK-1	IC_{50} for other PLKs.	Activity on Other Kinases	Clinical Development Stage (Clinical Trials ID)
BI2536[290]	ATP-competitive inhibitor	0.83 nM	PLK-2 =3.5nM PLK-3 = 9.0nM	IC_{50} >10μM for 63 kinases	Phase II
BI6727 (Volasertib)[303]	ATP-competitive inhibitor	0.87nM	PLK-2 = 5nM PLK-3 = 56nM	IC_{50} >10μM for over 50 kinases.	Phase III POLO-AML-2 (NCT01721876)
Compound 36[304]	Not determined	9.8nM	PLK-2 = 21nM PLK-3 = 178nM	High selectivity for PLKs over other kinases.	Preclinical
Cyclapolin 1[296]	Non-ATP competitive.	~20nM	Not determined	Src IC_{50} = ~100μM	Preclinical
DAP-81[297]	Predicted activity against the PLK-1 ATP-binding domain	0.9μM	Not determined	Not determined	Preclinical
GSK461364[305]	ATP-competitive inhibitor	2.2nM	PLK-2 >800nM PLK-3 >800nM	Some activity against KDR, FLT4, CSF1R and PDGFRB	Phase I[306]
HMN-176[307]	Disrupts PLK-1 cytoskeletal localisation	118nM	Not determined	Not determined	Phase I[308]
LFM-A13[309]	Kinase domain inhibitor	32.5μM	PLK-3= 61μM	BTK IC_{50} = 17.2μM	Preclinical
NMS-P937[310]	ATP-competitive inhibitor	2nM	PLK-2 = 10μM PLK-3 = 10μM	FLT3 IC_{50} = 510nM, MELK IC_{50} = 744nM CSNK2A1 IC_{50} = 826nM. 60 other kinases >10μM	Phase I ongoing (NCT01014429)
ON 01910.Na (Rigosertib)[295]	Affects microtubule dynamics. Phenotype may not be mediated via PLK1[297].	9nM but limited effect *in vitro* up to 30μM	PLK-2 = 260nM PLK-3 = >10μM	Multikinase activity - IC_{50}s; PDGFR = 18nm, Abl = 32nM, FLT = 42nM, Src = 155nM, Fyn = 182nM, CDK1 = 260nM	Phase II/III ONTRAC (NCT01360853)
Poloxin[298]	Inhibitor of PLK-1 PBD.	PBD = 4.8μM	PLK-2 PBD = 18.7μM PLK-2 PBD = 53.9μM	No reported non-PBD activity	Preclinical
Poloxipan[300]	Panspecific PBD inhibitor.	PBD = 3.2μM	PLK-2 PBD = 1.7μM PLK-2 PBD = 3.0μM	Weak activity against other phosphopeptide binding domains .	Preclinical
Purpurogallin[299]	PBD inhibitor	PBD = ~0.3μM	Some activity against PLK-2 PBD but not PLK-3	None reported	Preclinical
SBE13[311]	Blocks PLK-1 in an inactive state	Inactive PLK-1 = 0.2nM	PLK-2 = 66 μM PLK-3 = 875nM	Not Evaluated	Preclinical
ZK-thiazolidinone (TAL)[312]	ATP-competitive inhibitor	19nM	PLK-2 <100nM PLK-3 <100nM	High selectivity across a panel of 93 kinases.	Preclinical

Table 1-2 Published Clinical Trials of PLK-1 Inhibitors. Continued Overleaf

Compound	Phase	Trial Protocol	Number of Subjects	Cancer Types (n)	Endpoint	Reponses (RECIST criteria)	Treatment Related Toxicities (Most Frequent) or Deaths
BI2536	I[313]	Open label, dose escalation	40 patients	Advanced disease, multiple cancer types	Dose limiting toxicity (DLT)	2.5% partial response, 35% stable disease.	Neutropaenia (40%) Nausea (37.5%) Leukopaenia(35%) Anorexia (25%) Alopecia (20%)
BI2536	I[314]	Open label, dose escalation	70 patients	Advanced disease, multiple cancer types	DLT	No partial responses, 32% stable disease.	Neutropaenia (46%) Leukopaenia (40%) Anaemia (14%)
BI2536	I[315]	Open label, dose escalation	21 patients	Advanced disease, multiple cancer types	DLT	No partial responses, 38% stable disease.	Fatigue (62%) Leukopaenia (38%) Alopecia (33%) Constipation (29%) Neutropaenia (29%)
BI 2536	I[316]	Open label, dose escalation in combination with pemetrexed.	41 patients	Advanced or metastatic non-small cell lung cancer	DLT	5% partial response, 54% stable disease.	Fatigue (71%) Nausea (37%) Rash (34%)
BI 2536	II[317]	Open label, Day 1 (200mg infusion) or days 1-3 (50 or 60mg infusion)	95 patients randomised to either schedule	Stage IIIB/IV non-small cell lung cancer	Response & Progression free survival (PFS)	4% partial response. Median PFS 8 weeks.	Grade 4 neutropaenia (37%) Fatigue (31%) Nausea (27%) Deaths 1 patient.
BI2536	II[318]	Open label, single protocol multiple tumour types	71 patients	Head and Neck (n=14), Breast (n=14), Ovarian (n=15) Sarcoma (n=14), Melanoma (n=14)	Response & PFS	No partial responses, 30.5% stable disease. Median PFS 6 weeks.	Fatigue (39%) Alopecia (27%) Febrile neutropaenia (20%) Deaths 1 patient.
BI2536	II[319]	Open label, Day 1 (200mg infusion) or days 1-3 (60mg infusion)	86 patients randomised to either schedule	Advanced pancreatic adenocarcinoma	Response & PFS	2.3% partial response,. 24% stable disease. Median PFS 6 weeks	Neutropaenia (37%) Leukopaenia (29%) Fatigue (29%)
BI2536	II[320]	Open label, Day 1 (200mg infusion), dose escalation pending toxicity.	20 patients	Hormone refractory prostate cancer	Response & PFS	No partial responses, 40% stable disease.	Neutropaenia (73%)

Compound	Phase	Trial Protocol	Number of Subjects	Cancer Types (n)	Endpoint	Reponses (RECIST criteria)	Treatment Related Toxicities (Most Frequent) or Deaths
ON01910.Na	I[321]	Open label, dose escalation. Twice-weekly 2-hour i.v. infusion for three weeks.	20 patients	Advanced disease, multiple cancer types	DLT	5% partial response, 5% stable disease	Skeletal pain (30%) Fatigue (20%) Anaemia (15%)
ON01910.Na	I[322]	Open label, dose escalation. Once weekly 24 hour i.v. infusion	23 patients	Advanced disease, multiple cancer types	DLT	33% stable disease	Fatigue (48%)
ON01910.Na	I[323]	Open label, dose escalation. One 72 hour i.v. infusion.	27 patients	Advanced disease, multiple cancer types	DLT	15% stable disease	Fatigue (60%) Neutropaenia (40%)
ON01910.Na	I[324]	Open label, dose escalation. Twice-weekly 2-hour i.v. infusion for three weeks in combination with once weekly infusion of Gemcitabine	40 patients	Advanced solid tumours (n = 19) and 21 further patients with pancreatic ductal adenocarcinoma (PDAC)	DLT	Solid tumours; 10% partial response, 5% stable disease PDAC; 5% partial response, 58% stable disease	Death (2.5%) Neutropenia (27.5%) Thrombocytopaenia (15%) Lymphopaenia (10%)
ON01910.Na	I[325]	Open label, dose escalation.	14 patients	High risk myelodysplastic syndrome and acute myeloid leukaemia	DLT	43% partial Haematological or bone marrow response	Upper respiratory tract infection (28%) Haematuria (14%) Atrial fibrillation (14%)
ON01910.Na	II[326]	Randomised comparison of oral intermittent (2 out of three weeks) or continuous treatment	29 patients	Transfusion dependent myelodysplastic syndrome	Response & PFS	47% transfusion independent for >8 weeks	Urinary symptoms higher (60%) in continuous dosing arm vs intermittent (20%).
HMN-214	I[308]	Open label, dose escalation days 1-21 out of 28	33	Advanced disease, multiple cancer types	DLT	No partial responses, 24% stable disease	Fatigue (38%) Nausea (46%) Vomiting (46%)
HMN-214	I[327]	Open label, dose escalation days 1-5 out of 28	32	Advanced disease, multiple cancer types	DLT	3% partial response, 9% stable disease	Neutropaenia Neuropathy
BI6727	I[328]	Open label, dose escalation	65 patients	Advanced disease, multiple cancer types	DLT	5% partial response, 40% stable disease	Anaemia (22%) Neutropaenia (15%) Fatigue (15%)
GSK461364	I[306]	Open label, dose escalation	40 patients	Advanced disease, multiple cancer types	DLT	6 patients exhibited stable disease (15%). 4 of these had oesophageal cancer.	Neutropaenia (12.5%) Anaemia (10%) Thrombocytopaenia (10%) Thromboembolism (5%)

1.5.7 The Effect PLK-1 suppression on non-transformed cells.

Crosses between PLK-1 heterozygous mice do not yield viable homozygous knockouts[329]. Subsequent investigations reveal PLK-1 -/- embryos show spindle abnormalities and cannot develop beyond the 8-cell stage, confirming the essential role of PLK-1 in mammalian embryonic cell proliferation[329]. PLK-1 also appears to be ubiquitously expressed in proliferating human tissues[262,263].

It is surprising, therefore, that a divergent mitotic phenotype has been observed between HeLa cells and non-immortalised human foreskin fibroblasts in an early study of PLK-1 inhibition by antibody microinjection[239]. This prompted several studies comparing cancer cell lines and non-transformed cells. Viability was significantly reduced at 48 hours after PLK-1 expression was inhibited by siRNA in MiaPaCa and Hela cancer cell lines but not in human fibroblasts[281]. A small reduction in fibroblast viability was noted in PLK-1 siRNA treated cells at 96 hours post transfection suggesting the effects of PLK-1 inhibition may be evident in non-transformed cells in longer term assays. This was supported by a clonogenic assay which confirmed significantly reduced colony formation at 14 days after transfection[281]. Treatment of rat primary cardiac fibroblasts and Human umbilical Vein Endothelial Cells (HUVECs) with BI2536 also results in apoptosis[330].

In a further study, PLK-1 was suppressed by lenti-viral transduction of shRNA in cancer cell lines, the hTERT-immortalised retinal pigment epithelial cell line hTERT-RPE1, and the spontaneously immortalised breast epithelial cell line MCF10A. PLK-1 shRNA produced an accumulation of 4N cells and reduced viable cancer cell number but had no significant effect on cell cycle profile nor cell number in the non-transformed line, although the viability method used was not reported in this study[331]. Subsequent studies have confirmed hTERT-RPE1 cells can proliferate after reduction of PLK-1 expression by >90% using RNA interference methods but complete loss of the PLK-1 gene locus is associated with mitotic abnormalities[242].

Supporting this gene dosage effect, divergent cell cycle phenotypes have been observed with variable levels of PLK-1 suppression. In *Drosophila*, a partially active, weakly hypomorphic polo mutant results in persisting cell cycle progression

in contrast to inactivating strongly hypomorphic mutants which completely arrest proliferation[332]. In human cancer cell lines, partial PLK-1 knockdown leads to mitotic arrest and complete PLK-1 depletion leads to G2 arrest and failure to activate Cdc25c[333].

Raab and colleagues interrogated the role of PLK-1in normal adult mammalian tissues *in vivo* using a mouse model with doxycycline-inducible expression of PLK-1-targeting shRNA[334]. The non-transformed mouse fibroblast cell line NIH3T3 expressed 10-fold higher levels of PLK-1 than adult mouse GI-tract tissues, and other organs with low basal levels of cell proliferation but similar levels to mouse embryonic tissues, and adult mouse thymus and testis. After PLK-1 was suppressed for 6 weeks by doxycycline-treatment of adult mice, a comprehensive tissue analysis was performed. Spermatogenesis was found to be unperturbed by PLK-1 suppression, in contrast to findings in *Drosophila* after abrogation of the fly homolog *polo*[335]. Surprisingly, no evidence of marrow suppression was observed and control mice exhibited similar white cell and granulocyte counts to PLK-1 suppressed mice. This may be due to the residual 10-20% expression of PLK-1 being sufficient to maintain normal cell function. This could also be secondary to artificial dependence on PLK-1 in rapid proliferating cultured cells.

To test this, the authors used the same inducible expression system in primary mouse embryonic fibroblasts (MEFs) *in vitro* and found suppression of PLK-1 did not significantly reduce proliferation at 96 hours. In contrast a range of human cancer cell lines were found to be more sensitive to PLK-1 siRNA compared to HUVECS, human fibroblasts and keratinocytes.

The widespread observation of haematological toxicity in Phase I human clinical trials of PLK-1 inhibitors despite the apparent cancer-cell selectivity from preclinical studies suggests there may be important flaws in preclinical models. In particular the lack of marrow suppression in the conditional PLK-1 shRNA mouse model is concerning for future toxicity prediction and translation of preclinical therapeutic leads from rodent to man. Importantly, MEFs treated with BI2536 showed significantly reduced proliferation and induction of apoptosis at doses >100 nM[334].

This suggests PLK-1 suppression by RNA interference may not be a suitable method to model the phenotype of PLK-1 inhibition using kinase-inhibitors.

1.5.8 Enhancing specificity through synthetic interactions.

Identifying the features of cancer cells which confer sensitivity to PLK-1 inhibitors may allow the therapeutic window between malignant and non-transformed cells to be widened. Somatic mutations are commonplace in cancers and may disrupt signalling networks to induce the transformed phenotype. The same mutations may also create a unique dependency on a signalling pathway that is not present in normal cells. Inhibiting this pathway would therefore only produce significant effects in cells with the appropriate mutations. One commonly mutated oncogene in solid tumours is the inhibitory GTPase KRAS. Mutations commonly lead to a loss of GTPase activity either directly, or by preventing the binding of GTPase-Activating Proteins, leading to constitutively active KRAS signalling[336].

A search for synthetic lethal interactions with the oncogenic KRAS G13D allele has been conducted using a genome-wide RNA interference strategy. For this study, two isogenic colorectal adenocarcinoma cell lines bearing either KRAS wt/G13D alleles or KRAS wt/- were transfected with pools of shRNA sequences[337]. Sequences that were significantly depleted in KRAS wt/G13D cells over subsequent population doublings were identified and a subset subsequently validated on a candidate basis. A number of genes involved in mitosis were proposed as synthetic lethal candidates, with PLK-1 identified as the only kinase in this study. PLK-1 suppression using shRNA significantly depleted cells with KRAS wt/G13D compared to KRAS wt/- and KRAS wt/G13D cells were also sensitive to BI2536 *in vitro* and in a nude mouse xenograft assay[337]. The possibility to preferentially target cells with activating KRAS mutations provides a compelling rationale for further development of the PLK-1 inhibitor class and highlights the importance of identifying predictors of sensitivity.

1.5.9 Cross regulation between PLK-1 and p53

The high frequency of TP53 mutation in cancer also makes this gene a compelling target for a synthetic lethal strategy. The interaction of p53 and PLK-1 is complex.

When p53 was co-suppressed with PLK-1 using a lentiviral shRNA vector, proliferation was inhibited in MCF10A cells suggesting wild-type p53 protects cells from the effects of PLK-1 depletion[331]. Depletion of PLK-1 produced a greater reduction in viable cell number in cell lines with wild-type p53 inhibited by the E6 oncoprotein[198] or SV40 large T antigen[173] compared to identical cells with uninhibited p53[281]. Furthermore, inhibition of the DNA damage sensor ATM has also been reported to enhance the lethality of PLK-1 depletion, potentially by blunting p53 activation[283]. These data suggest short-term inhibition of PLK-1 may preferentially target cancer cell lines and loss of p53 function may potentiate this effect.

Hela cells depleted of PLK-1 or treated with PLK-1 inhibitors show stabilisation of p53[282,338]. Temporary loss of p53 expression mediated by siRNA does not appear to potentiate the PLK-1 effect suggesting a more prolonged inhibition is required[281].

The activity of the PLK-1 promoter can also be indirectly supressed by p53. In response to genotoxic stress, p53 regulates expression through transcriptional repression rather than promoter site binding to mediate the apoptotic cascade[339]. Wild-type p53 represses PLK-1 expression and regulates PLK-1 expression during the G2/M transition[340,341]. This has been proposed to be mediated by p21-dependent[342] and independent mechanisms[340].

PLK-1 can also directly bind p53 within the DNA-binding domain resulting in reduced p53-mediated transcriptional activity in response to DNA damage[343]. It remains to be conclusively determined whether PLK-1 phosphorylates a specific residue on p53 to mediate this effect[344]. PLK-1 can also indirectly regulate p53 protein levels via phosphorylation and activation of the p53-regulating E3 ubiquitin ligases, MDM2[345] and Topors[344].

This reciprocal negative regulation suggests a model where loss of p53 function by mutation in many cancers allows PLK-1 over-expression and inhibition of PLK-1 induces wild-type p53 expression. This is supported by one study in breast cancer tissue where mutation of p53 was significantly correlated with high expression of

PLK-1. Notably, the combination of p53 mutation and PLK-1 overexpression defined a group with significantly poorer prognosis than either p53 mutation or PLK-1 overexpression alone[264].

1.5.10 The status of p53 as a determinant of PLK-1 inhibitor sensitivity

In a study using isogenic HCT116 cells the dose- responses to BI2536 under basal growth conditions were independent of p53 status. A follow-up study from a different group used the same cell lines and concluded that p53 function does not determine response to PLK-1 inhibition. A similar induction of apoptosis and reduction of proliferation was noted on transfection of siRNA to PLK-1 in both wt and p53 null cells[338]. Very similar results were also noted with transfection of scrambled siRNA suggesting that the cytotoxicity during transfection may mask the results of PLK-1 suppression[338].

In the same study, the responses to the PLK-1 inhibitors BI2536, BI6727 and Poloxin were examined and minimal differences in proliferation rates noted between cells isogenic for p53. Unfortunately IC_{50} values were not calculated to allow comparison with other studies. The inhibitors also appeared to produce similar mitotic phenotypes in the isogenic cells[338]. Retention of p21 induction after DNA damage has been observed in the p53 null HCT116 cells, potentially through a p65-mediated mechanism[346,347]. This does not appear to explain the cell cycle effects since p21 was not induced in HCT116 p53 -/- cells after PLK-1 inhibitor treatment[338]. MCF7 and A549 cells both exhibited unchanged responses to PLK-1 inhibitors when wild-type p53 was suppressed by siRNA[338]. The lack of an augmented phenotype from PLK-1 inhibition when p53 expression is temporarily reduced has been previously observed[281]. Low levels of residual p53 may be sufficient to mediate the effect in this case.

The aneuploid human embryonic kidney cell line HEK 293 exhibits G2/M blockade after PLK-1 suppression despite wild-type p53 function suggesting p53 status is not

the only determinant of susceptibility to PLK-1 inhibitors[281]. The cross inhibition of other polo-like kinases by BI2536 may also confound comparisons with RNA interference approaches and other compounds with greater specificity for PLK-1 over other pololike kinases.

Colony outgrowth was evaluated after a 72 hour treatment with the PLK-1 inhibitor GSK461364A across a panel of 43 cell lines representing various cancer types. Gene expression, copy number variation and p53 genotype were correlated with sensitivity. Mutation of p53 was 88% sensitive and 67% specific for the prediction of sensitivity to this compound[348]. Those sensitive cell lines with wt p53 were also found to express higher levels of MDM2 compared to resistant cells with the wild-type allele or sensitive cells with p53 mutations[348]. Interestingly, neither PLK-1 mRNA expression, nor KRAS or BRAF mutation status correlated with sensitivity to the inhibitor used in this study[348].

The apparent therapeutic index observed between wild-type and p53 mutant cell lines may be a consequence of the assay used. Studies with GSK461364A in a standard 72 hours proliferation assay identified differences between cell lines that correlated with proliferative rate rather than p53 genotype[349]. The effect of p53 on colony outgrowth after PLK-1 inhibitor treatment has not yet been assessed using isogenic cells. This would help to clarify the effect of p53 genotype on response to these compounds.

1.5.11 Strategies to augmenting sensitivity to PLK-1 inhibitors

The rationale for the use of PLK-1 inhibitors in cancer is that they target rapidly proliferating cells. If normal cells can be induced to underdo temporary cell cycle arrest, they may be spared from the toxic effects of PLK-1 inhibition. Activation of wild-type p53 by inhibition of MDM2 has been reported to lead to cell cycle arrest but not cell death in non-transformed cells through activation of the G1/S and G2/M checkpoints[350-352]. Tumour cells with mutant p53 have been reported to be insensitive to this treatment and continue to progress through the cell cycle.

Combining an activator of wild-type (wt) p53 with a cytotoxic cell cycle inhibitor, referred to as "cyclotherapy", represents a potential strategy to enhance the therapeutic window between cancer and normal cells.

A reportedly specific inhibitor of the MDM2-p53 interaction, Nutlin-3A has been employed for this purpose. When p53 was activated by gamma irradiation or Nutlin-3A treatment, HCT116 cells with wt p53 were resistant to the effects of BI2536 in a short term viability assay relative to HCT 116 cells without p53[220]. Interestingly, in the same study, Nutlin-3A significantly reduced viability in all cell lines at 10 μM regardless of p53 genotype and p53 null cells were only ten to fifty-fold more sensitive than cells with wt p53. Treatment with BI2536 eradicated established HCT116 p53 -/- xenografts in nude mice and Nutlin-3A treatment did not affect tumour sensitivity to BI2536 but did reduce the BI2536-induced neutropaenia[220]. Nutlin-3A treatment alone also appeared to reduce HCT116 p53 -/- tumour growth rates compared to control and Nutlin-3A treatment could not protect HCT116 wt p53 cells from the cytotoxic effect of BI2536.

This suggests that Nutlin-3A may have undesirable off-target effects and p53 activation using this compound may not be sufficient to protect against the effects of PLK-1 inhibitor treatment. Cyclotherapy relies on non-genotoxic activation of p53 in normal cells to induce G1/S arrest. A recent report has highlighted a further undesirable toxicity of Nutlin-3A treatment with the observation of *de novo* p53 mutations arising in cells with previously wt p53 treated with Nutlin-3A[353].

To identify synthetic lethal interactions with the PLK-1 inhibitor GSK461364, an RNAi screening approach was employed using a barcoded library of 4603 shRNAs directed against 1657 genes. A group of differentially enriched shRNAs corresponding to 97 genes was identified[354]. Retinoic acid receptor inhibition was found to induce resistance to the inhibitor. Conversely a combination of retinoic acid receptor agonists was found to enhance sensitivity by an average of 10 fold[354]. Further studies are awaited to confirm if this is a drug specific effect or applicable to all PLK-1 inhibitors.

In a phase I study of the PLK-1 inhibitor GSK461364, 4 out of the 6 patients exhibiting stable disease had oesophageal cancer and one patient with metastatic OAC exhibited stable disease for over 1 year[306]. Although very preliminary data, this apparent relative sensitivity may reflect a disease-specific vulnerability to this class of inhibitor. This may be a consequence of a high natural frequency of TP53 mutation in this cancer type, or other undetermined genetic susceptibility.

1.5.12 Targeting PLK-1 in Oesophageal Cancer

Therapeutic targeting of PLK-1 has only been reported to date in squamous cell carcinomas of the oesophagus. In one study, partial PLK-1 suppression by shRNA transfection reduced, but did not abolish, proliferation of OSCC cells and induced caspase-3 activation and apoptosis[269]. A further study in OSCC cell lines confirmed reduction of proliferation and viability with induction of apoptosis 48-72 hours after transfection of siRNA to PLK-1[355]. Colony formation and xenograft formation has also been shown to be reduced after transient transfection of siRNA to PLK-1 in OSCC cells[271,272].

PLK-1 inhibitors have also been investigated and complete regression of established xenografts of the OSCC cell line KYSE510 was observed when twice weekly tail vein injections of 50 mg/kg BI2536 were administered[356]. This more robust assay confirms a tumour cell killing effect rather than a cytostatic phenomenon *in vivo*.

A study into the regulation of PLK-1 in oesophageal cancer cell lines identified hsa-miR-539-5p as a potential endogenous regulator of PLK-1 expression[270]. A complementation site in the PLK-1 3'-UTR was identified and, importantly, transfection of hsa-miR-539-5p reduced both PLK-1 expression and proliferation in cell lines[270]. The activity of this mechanism was not assessed in non-transformed cells and the relevance of this mechanism *in vivo* is unknown. Small molecule drug-like mimetics of this microRNA may however be useful as alternative strategies to target PLK-1.

PLK-1 has also been shown to enhance the growth of OSCC cells in suspension, a key property of metastasising cells *in vivo*[357]. This was dependent on an active PLK-

1 kinase domain which prevented the proteasomal degradation of β-catenin. The transcription factor RelA, a component of the NF-κB complex, was subsequently proposed as an upstream regulator of PLK-1 expression in detached OSCC cells. This transcription factor complex is widely implicated in the link between inflammation and cancer and may represent a therapeutic target in its own right[358].

A follow-up study by the same group confirmed STAT3 was frequently constitutively active in OSCC cells and tissue, compared to normal squamous tissue. They also identified STAT3 positively regulates PLK-1 expression, potentially via a STAT3 consensus sequence in the PLK-1 promoter, and that PLK-1 could augment STAT3 expression[356]. This reciprocal activation enhanced cell survival *in vitro* and when cells were treated with BI2536, relative levels of active STAT3 decreased. Interestingly, constitutively active STAT3 expression partially protected KYSE510 cells from the apoptosis induced by BI2536[356]. Although most of the data was produced in one OSCC cell line, this study provides a clear rationale for the assessment of combination therapy with STAT3 and PLK-1 inhibitors.

1.6 Immunotherapy – an alternative therapeutic strategy.

The human immune system has evolved over millions of years to cope with a huge diversity of pathogens. The remarkable sensitivity of antigen recognition mechanisms allows both non-human tissues to be recognised from diverse evolutionary kingdoms but also allows the clearance of host cells that have developed mutations[359]. A hallmark of established cancers is that they manage to evade recognition by the immune system[360]. Reactivation or augmentation of the immune response to cancer therefore provides a compelling therapeutic strategy.

1.6.1 Mechanisms of immunotherapeutic activity

The most commonly method of cancer immunotherapy uses antibodies directed against cell-surface antigens expressed on tumour cells. Rituximab, trastuzumab, cetuximab have all reached clinical practice as targeted agents direct against CD20, HER2 and EGFR respectively[359].

The mechanisms of action for therapeutic antibodies are diverse and context dependent for each agent. Anti-tumour activity can be achieved directly by blocking growth promoting ligand-receptor interactions as in the case of EGFR-family directed antibodies. Much research has been focussed on expanding this paradigm to other cancer types but the mechanistic specificity is vulnerable to the development of downstream resistance[361,362].

The presentation of the crystallisable fragment (Fc) portion of IgG on the surface of cancer cells can also play an important role in engaging innate immune effector mechanisms via Fc receptors. This important mechanism of activity can lead to immune cell activation and cell-mediated cytotoxicity (antibody-dependent cellular cytotoxicity; ADCC)[363]. The specific mechanisms of ADCC *in vivo* remain an area of intense investigation but appear to include both direct phagocytosis by stimulated macrophages, neutrophils and other phagocytes and the development of acquired immunity via the induction of tumour-directed cytotoxic T-cells and host-derived anti-tumour antibodies[364].

1.6.2 T-cell Meditated Therapies

Improvements in the understanding of the microenvironment of the cancer niche has led to interest in T-cell based therapies. Modulating the cytotoxic T-cell response to cancer has been attempted using several strategies. CD4 positive T-helper cells regulate cytotoxic CD8 positive T-cells through combination of humoral and direct mechanisms including competition for CD80 and CD86 binding on antigen-presenting cells, preventing cytotoxic T cell activation[365]. Cytotoxic lymphocyte-related antigen-4 (CTLA-4) is present on CD4 positive cells and a monoclonal antibody to this immunosuppressive molecule, Ipilimumab, enhances cytotoxic T-cell activation[366]. Crucially, this antibody enhances cancer-specific survival in metastatic melanoma and is being investigated in several other cancer types[367]. Significant side-effects of this therapy, however, include various autoimmune effects including skin rash, colitis and hepatotoxicity.

A further approach includes the isolation of tumour-associated T cells directly from cancer biopsies, expansion and activation of these cells *in vitro* and then reinfusion

after systemic immunosuppression by lympho-depletion with cytotoxic chemotherapy with or without radiotherapy. This method has been shown to produce response rates of up to 70% of patients with metastatic melanoma[368].

In a variation of this method, cytotoxic T-cells are engineered to express T-cell receptors with high affinity to putative tumour cell specific antigens. This adoptive cell transfer method has been attempted for the melanocyte differentiation antigens gp100 and melanoma antigen recognised by T-cells1; (MART1) and clinical trials resulted in mild to moderate melanoma regression but also toxicity to normal melanocytes in the skin and eye.

In a further study using T-cells engineered to express a high affinity receptor to the melanoma antigen family A protein 3; (MAGEA3), five out of nine patients experienced objective tumour responses[369]. Unexpectedly, four patients developed severe neurological toxicity and subsequent investigations determined T-cell receptor cross-reactivity with MAGEA12 and MAGEA9, both of which were expressed at low levels in the brain. A further trial with engineered T-cells expressing a different receptor targeting MAGEA3 resulted in fatal cardiac toxicity due to cross-reactivity with the myocardial protein titin[370]. Serious toxicities from off-target effects have also been observed with T-cells directed against the carcinoembryonic antigen (CEA) expressed in colon cancer and HER2[371,372].

These studies confirm the significant *in vivo* efficacy of cytotoxic T-cell therapies but highlight the formidable potential toxicities associated with infusing activated T-cells with high-affinity towards epitopes not exclusively expressed on cancer cells. Indeed the success of most immunotherapeutic strategies relies on targeting a cancer specific antigen. Identifying these antigens is therefore a critical step in the development of these therapies.

1.6.3 Cancer-cell Antigens

Strategies directed toward identifying candidate cell-specific antigens for antibody-based therapeutics have previously emphasised the selection of cell-surface or secreted proteins. This follows the dogma that immunoglobulins are too large to

access intracellular epitopes. An intriguing recent study has demonstrated antibodies directed against intracellular proteins can arrest the development of spontaneous tumours and xenografts in immune-competent mice[373]. One possible explanation for the efficacy of antibodies directed to putatively intracellular targets is that proteins are actually available on the cell surface. This phenomenon has been observed with chaperones normally confined to the endoplasmic reticulum and cytoplasm in non-transformed cells identified on the cell membrane in cancer cells[374]. Nevertheless, if this paradigm holds true in humans, the range of candidate cell-specific targets would be significantly broadened[373].

Tissue specific proteins can also be used as targets for novel antibody-based imaging agents to assist in clinical staging[375] or, in the case of secreted proteins, for blood biomarker assays. Each of these applications could have clear clinical impact in OAC but no specific markers have yet been identified. To identify these candidate proteins *de novo*, expression must be measured using unbiased proteomic methods.

1.7 Proteomics

1.7.1 The Challenge of Measuring the Proteome

The study of the entire protein content of an organism, tissue or cell was first described as proteomics nearly twenty years ago[376]. In that time frame, the goal of measuring an entire eukaryotic proteome has been achieved[377]. Despite significant methodological advances, however, the human proteome remains to be measured *in toto* almost a decade after the publication of the complete human genome[378]. Several reasons can be proposed for this disparity.

The polymerase chain reaction (PCR), allows template nucleotide sequences to be copied with an increase in number of many orders of magnitude with very low error rates[379]. Complementary-base pairing also allows cryptic nucleotide sequences to be rapidly deciphered[380]. The combination of these methods and advances in computational assembly of short sequence reads allows nucleotide sequencing to proceed in massively parallel configurations to produce entire genomes within hours[381].

In contrast, the *de novo* identification of protein sequences contains greater intrinsic challenges. No method exists to amplify protein or peptide sequences and therefore proteomic methods are always restricted by the input mass. Similarly amino acids do not exhibit complementation and identification relies on mass measurement or, historically, chromatography or electrophoresis[382]. The proteome is also significantly larger than the genome with alternative splicing and alternative transcription start sites contributing to transcriptome and ultimately proteome diversity[383].

A further challenge is posed by greater combinatorial possibilities with up to 21 amino acids used interchangeably to generate peptides. This complexity is further increased by post-translational modifications (PTMs) including the addition of biochemical groups such as a phosphate (phosphorylation), a carbohydrate (glycosylation), and at least 25 other distinct moieties or modifications[382].

A final compounding difficulty is the dynamic nature of the proteome. The germline genome sequence of an organism is constant across all cells in that organism and is relatively stable in the face of DNA extraction methods even allowing DNA sequences to be obtained from ancient specimens[384]. In contrast the proteome varies from cell to cell and is highly context dependent with the post-translational state of a single protein varying across subcellular localisations. Extracting the proteome for quantification is also confounded by the rapid alterations in the PTM state induced by hypoxia and changes in intracellular pH with some phosphorylations reported to be lost within 60 minutes of tissue biopsy[385]. Many of these challenges have been addressed with recent technological advances, the most significant of which is the use of high accuracy mass spectrometry[386].

1.7.2 Mass Spectrometry for Proteomics

Mass spectrometry (MS) has become the definitive method to identify and quantify proteins *de novo* from complex mixtures. In essence, a mass spectrometer consists of a detector coupled to a mass analyser that measures both the number and the mass to charge ratio (m/z) of ions generated into the gas phase from an ionisation source. Variations on the instrumentation abound however, each with their own strengths

and weaknesses. A significant technological advance was made with the development of the Orbitrap mass spectrometer[387]. These instruments combine high sensitivity, high mass accuracy and cover wide m/z ranges to finally bring the measurement of whole proteomes within reach.

Electrospray ionisation (ESI) sources take a liquid, commonly a polar volatile solvent eluted from a chromatography column, and ionise the analytes directly into the gas phase[388]. These sources are most commonly used for the analysis of complex mixtures including cell lysates. Alternatives include matrix-assisted laser desorption ionisation (MALDI) sources which use a laser to ionise analytes directly into the gas phase out of a solid matrix[389]. These sources are limited in the number of ions that can be generated and have previously been reserved for relatively homogenous analytes.

1.7.3 Protein Identification using MS – Bottom-up Approach

The *de novo* identification of proteins from a complex mixture can be achieved by several means. The most common method, termed "shotgun" or "bottom-up" proteomics relies on the identification of peptides generated by proteolytic digestion of the protein mixture. The presence of a protein in the original mixture is then inferred by interrogation of a protein sequence database with the identified peptide sequences. Matching a peptide sequence unique to a particular protein provides evidence of the protein in the original mixture[386]. An example workflow is illustrated in Figure 1-6.

Shotgun proteomics relies on tandem mass spectrometry (MS/MS) where peptides are ionised to generate precursor ions, analysed and separated according to their m/z in the primary mass spectrometry run (MS1). Precursor ions are then selected for fragmentation, usually by collision ion dissociation, and the fragment ions are separated and analysed in the second MS run (MS2)[390]. Multiple fragment species are generated from the same peptide and, with high quality spectra and sufficient fragment ions, species differing by each individual amino acid in the peptide will be discernible as discrete ion peaks separated by a measured mass difference. As amino acids all have a fixed, defined mass, the measured difference can be used to identify

the amino acid. Thus the sequence of the peptide can be determined directly; defined as *de novo* peptide sequencing[391]. In practice, with complex peptide mixtures it is rarely possible to sequence all peptides directly and this labour intensive approach is reserved for organisms with limited genome sequence information and therefore limited or absent potential protein databases.

More commonly, database searching is performed to generate peptide-spectrum matches. A number of algorithms have been described but in general the measured precursor mass is used to filter a database of peptides generated by *in silico* digestion of a list of potentially identifiable proteins. Theoretical fragment ion mass differences are generated for all the candidate peptides with a matching precursor mass. These are compared with the identified fragment ion spectra and candidates ranked using a scoring algorithm, specific to the database search method[392].

These methods identify peptides without the requirement for prior mass spectrometry. Recently, organism-specific spectral libraries generated using stringent identification thresholds and evaluation of millions of published experimentally-derived peptide spectra have become available[393]. An alternative, or complementary, approach is to search identified spectra against these libraries, incorporating other spectral features such as relative ion intensity. This has been reported to enhance the number of peptide identifications compared to standard database search strategies[393,394]. These algorithms may become more widespread in the future as spectral libraries become more complete but this approach has not yet been widely adopted.

Peptide sequences are then used to identify proteins from the original search database. A variety of statistical approaches are included in commonly used software packages to deal with protein inference problems such as repeated peptide sequencing events, peptides shared between multiple proteins and estimating the false-discovery rate[395].

A significant limitation of mass spectrometry is the throughput of ions that can be analysed. Although this has improved with the current generation of Orbitrap

instruments[396], the number of analytes that can be studied simultaneously without loss of data is still restricted. Tissue lysates contain highly diverse mixtures of proteins. This diversity is further compounded by proteolytic digestion, presenting significant challenges for peptide spectrum matching. To reduce the complexity of the original sample, fractionation approaches are often employed. These can include strong cation exchange[397], subcellular fractionation[398], isoelectric focusing electrophoresis[399], and other chromatography methods[400]. By delivering fractions with reduced numbers of unique peptides into the mass spectrometer, homogenous m/z fractions can be produced during the MS1 phase which can be accurately sequenced during the MS2 phase[401,402].

1.7.4 Data-Dependent and Independent Shotgun Proteomics

A key feature of the shotgun proteomic method as described previously is the selection of precursor ions for fragmentation in the MS2 phase. This is usually performed on the basis of precursor ion intensity and is referred to as a data-dependent approach[403]. This method has the limitation that a precursor ion must be detected to allow peptide sequencing and places an intrinsic bias towards abundant precursor species. An alternative strategy, exemplified by the Precursor-acquisition Count Independent from Ion Count (PAcIFIC) method, is to fragment all precursor ions within defined m/z windows regardless of whether a precursor ion was detected or not[404]. The precursor mass used for peptide spectrum matching is assigned as the centre of the MS1 m/z window. When this method is applied, fragment ions yielding high-confidence peptide spectrum matches can be detected in up to 10% of cases in the absence of a precursor ion[404,405]. Applied systematically across a wide range of precursor m/z windows, this approach can enhance the depth of proteome coverage by identifying low abundance peptides. A disadvantage is the long data acquisition times required to obtain spectra across all m/z windows. With optimised liquid chromatography gradients, however, the mass spectrometry time can be reduced by over 50%[406].

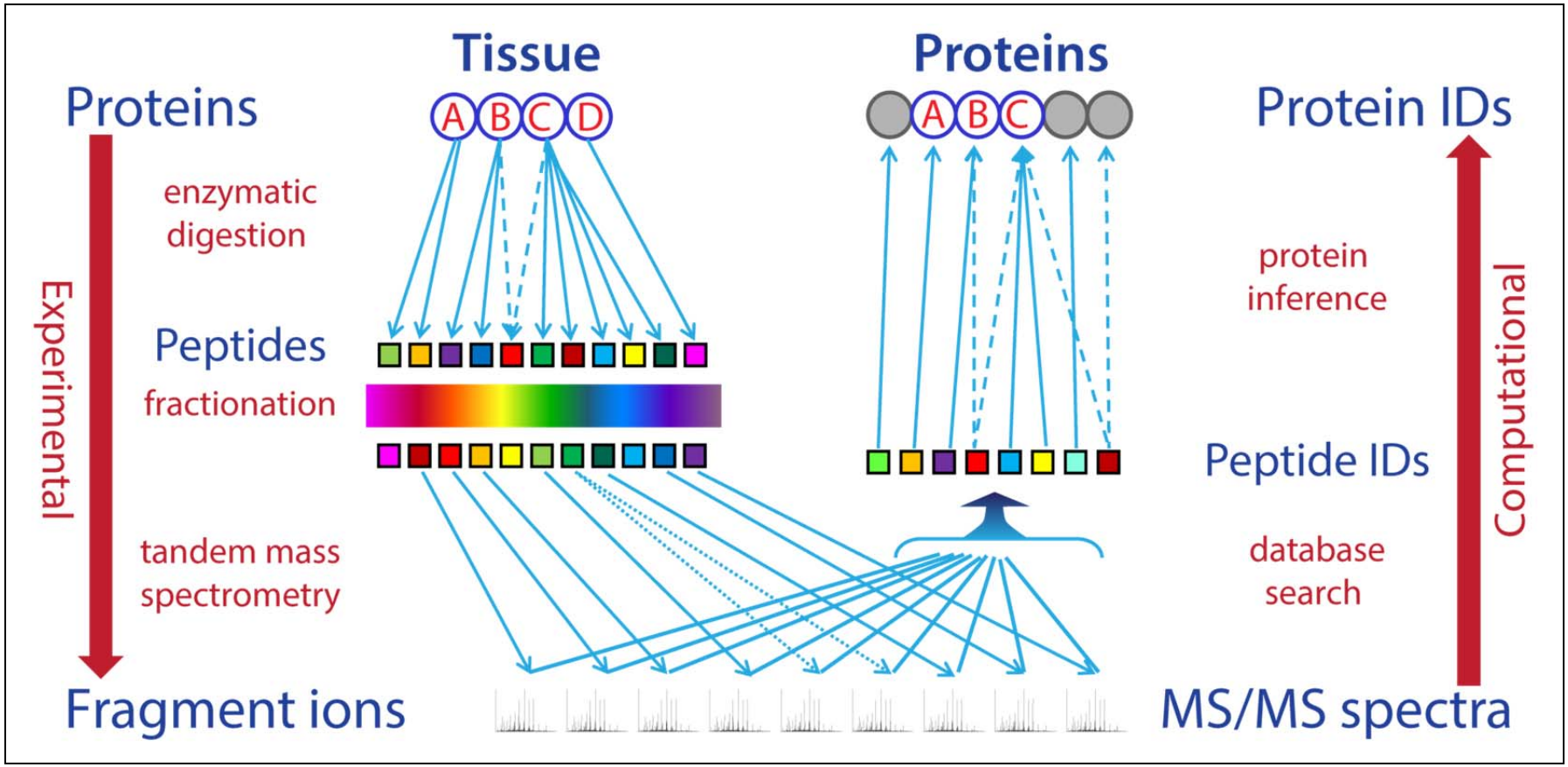

Figure 1-6 An Overview of Protein Identification by Shotgun Proteomics. A complex protein mixture, in this example a tissue sample containing proteins A-D, is proteolytically digested to yield peptides. Each peptide is schematically represented as a coloured box. To reduce the mixture complexity, peptides are fractionated by a common property. In this illustration the property is colour but in practice this is a biochemical property such as isoelectric point. Peptide fractions are subjected to tandem mass spectrometry to yield fragment ion spectra. Peptide-spectrum matches (IDs) are made using a protein database, the peptide (precursor) ion masses and a database search tool[407]. Not all fragment ion spectra result in a peptide match and some peptide matches are of low confidence (e.g. light green peptide). Using further statistical tools[395], proteins are identified with unique peptide matches confirming the presence of a protein in the original mixture. Each shotgun experiment only identifies a subset of the proteome from complex mixtures such as tissue lysate so in this example protein "D" has not been identified.

1.7.5 Protein Identification using MS – Top-Down Approach

An exciting recent development has been the enhancement of techniques to identify intact proteins by mass-spectrometry; a so-called "top-down" approach. In a recent report, over 1000 distinct protein isoforms could be identified using cultured mammalian cells and extensive orthogonal fractionation in the liquid-phase[408]. An advantage of this method is the direct identification of proteins, rather than inference from peptide identifications using the shotgun approach. This provides the potential to characterize the entire population of isoforms generated from a single gene and identify dynamic changes in protein-processing, alternative-splicing or post-translational modification often not possible from peptide-level data alone. Although not currently capable of proteome-scale analysis, with further developments in automated fractionation, instrumentation and data-analysis methods, this may become feasible in the future[409]. Measuring dynamic changes in cellular states, the ultimate goal of most biological proteomic experiments, however, requires quantitation in addition to protein identification and methods to undertake this using a hypothesis-free top-down approach remain to be described.

1.7.6 Gel-based Methods for Quantitative Proteomics

A typical proteomics experimental design would be to compare a cell line under two or more conditions and attempt to identify differentially expressed proteins. Historically, 2D gel electrophoresis would be used to separate the cell lysates from each condition according to protein mass and isoelectric point[410]. Gels could then be stained using a silver-based or other similar methods and differentially expressed proteins could be identified as spots of differing intensities[411].

A variation of this method minimised the gel to gel variability by labelling all the proteins in each sample with a different fluorophore and running all the samples together on the same gel[412]. By quantifying the relative emission from each fluorophore across the spots, the relative expression could be determined.

In both examples, protein spots were excised from the gels, digested to peptides using proteolytic enzymes and subjected to mass spectrometry for peptide and subsequent protein identification using similar strategies to shotgun proteomics. This

method has the advantage of limiting the protein identifications to a small number of differentially expressed proteins, and providing a relatively homogenous sample for mass spectrometry. Unfortunately, despite advances in the automation of spot detection and quantification, these methods were only semi-quantitative, labour-intensive, the data quality was highly user dependent and protein identifications were limited to a few dozen per experiment.

1.7.7 MALDI-Imaging MS (MALDI-IMS)

A major disadvantage of lysing tissue biopsies for down-stream mass spectrometry analysis is the loss of microscopic spatial information relating to protein expression. Epithelial-derived tumour cells exist, and in many cases are dependent on, the interactions and cellular context of their microenvironment[360]. Understanding the changes in protein expression that occur within epithelial and stromal tumour compartments may unveil novel insights not apparent from analysis of the tumour biopsy *in toto*. Similarly, cell types of one origin exhibit varied expression profiles across tumours reflecting mutational diversity and changes in oxygen tension and extracellular nutrient availability[413].

To capture this heterogeneity, MALDI techniques have been adapted to allow direct ionization and mass-spectrometry from tissue sections[414]. By co-registering spectra and histological images, patterns of protein expression can be interpreted within their biological and cellular context. The significant advantages of this imaging mass spectrometry method are offset by some of the limitations of other MALDI approaches.

MALDI-IMS generates spectral features (m/z) which can be used to differentiate samples but does not identify proteins directly. Hybrid approaches with downstream tandem mass spectrometry allow low mass proteins to be identified directly although proteome coverage has not yet reached parity with LC-MS/MS analysis of tissue lysates[415]. A further significant limitation is in the resolution of ionization sources. Current technologies allow a minimum resolvable area of 10 µm but most analyses are practically limited to areas of 100 µm[416]. This allows a granular expression map

to be generated but the goal of identifying subcellular expression patterns, for example at the tumour-stromal interface, remains elusive.

1.7.8 Quantitative Shotgun Proteomics

Advances in sample processing and instrumentation have enabled the development of quantitative shotgun proteomic methods. These rely on lysis, digestion and usually fractionation of samples prior to liquid chromatography (LC) and MS/MS. A labelling phase can be incorporated into the sample preparation stages prior to MS/MS or peptides can be quantified directly using label-free strategies[417].

1.7.9 Quantitative Shotgun Proteomics by Labelling

1.7.9.1 Stable Isotope Labelling of Amino Acids in Culture (SILAC)

Chemical labelling can take place at the protein or peptide level. The use of stable carbon, hydrogen and nitrogen isotopes have allowed differential labelling of amino acids that by their mass differences are resolvable as discrete spectral peaks. This approach, termed Stable Isotopic Labelling of Amino Acids in Culture (SILAC), allows the proteins in mammalian cells in culture to be isotopically labelled by the use of medium containing only "heavy" amino acids such as Leucine, Lysine and Arginine[418]. A typical experiment would comprise one treated, "heavy"-labelled cell line and a control unlabeled, "light", cell line. Cell lysates would be mixed in a 1:1 ratio and then subjected to standard LC-MS/MS workflow. Peptides would be identified in the usual fashion and the relative expression between cell line conditions identified at the MS1 level by the ratio of heavy to light peptide ion intensities. This approach has been shown to be reproducible across the proteome with relative standard deviations of protein ratios of ~30%[417]. By using both heavy lysine and heavy arginine combinations, three conditions can be compared simultaneously.

A disadvantage of SILAC approaches is the requirement for complete label uptake by cultured cells, which limits the application to cells which express stable phenotypes of interest across several passages. The requirement for prior labeling in the conventional SILAC method also precludes the study of human tissues samples

although fully isotopically-labelled organisms have been described which may have application in disease models[419-421].

1.7.9.2 Super-SILAC

A recently described variation of the SILAC method, termed super-SILAC, has potential application for the quantitative proteomic study of human cancer samples[422]. In this procedure a mixture of cell lines derived from the cancer tissue of interest and approximately covering the expression profile of the tissue of interest are heavy-labelled using the SILAC method. A mixture of lysates from these cell lines with a defined protein mass is spiked-in to each tissue lysate in a 1:1 ratio before digestion, fractionation and LC-MS/MS using standard procedures. Peptide identification and quantitation then proceeds as for a standard SILAC experiment. The ratio of expression between heavy and light peptides is calculated for each tissue sample. The constant SILAC spike-in mass provides a method of normalizing between experimental runs and also, by calculating the ratio of ratios, allows the relative expression between tissue types to be calculated[423]. An advantage is the spike-in standard can be used in multiple experiments on multiple platforms and still allow normalization between experiments and, once the spike-in standard is generated, there is no further labelling steps or reagent costs. A disadvantage is that the accuracy of SILAC is highest at ratios <2 and therefore a relatively close match to the tissue expression profile is required[417]. SILAC media is not yet available for many primary cells and therefore primary human tissues or cancers with few available cell lines may be difficult to analyse with this technique. Similarly, proteins specific to a particular tissue sample will not be quantified.

1.7.9.3 Isotope-coded Affinity Tags (ICAT)

In this method, the cysteine residues of reduced proteins are labelled with tags comprising a composite of a sulfhydryl reacting group, a deuterated linker, and a biotin affinity tag[424]. Proteins from discrete samples can be differentially labelled as both “light” and “heavy” isotopes of the linker are available. Labelled samples are then pooled and digested together. Cysteine containing peptides are then enriched by avidin-affinity chromatography. Peptides can then be further fractionated or directly

subjected to LC-MS/MS. The different isotopes of the deuterated linker provide discrete mass peaks during MS1 analysis to allow differential expression analysis.

Unfortunately only cysteine containing proteins can be studied, limiting proteome-wide efforts and the bulky affinity group, biotin, introduces significant background into the MS/MS spectra[425]. Furthermore, deuterated labels are more hydrophobic and therefore are differentially eluted during reverse phase LC, complicating the MS analysis[426]. This technique still has a role, however, as the affinity enrichment step allows the study of low abundance proteins, not easily accessible by other methods.

1.7.9.4 ^{18}O labelling

Accurate quantitation can also be achieved by labeling at the peptide level. An advantage of this strategy is that it can be applied to almost any sample. In an approach that predates the SILAC method, samples for comparison are either proteolytically digested in ^{18}O-containing water for the "heavy" sample or standard "light" water[427]. As the protease, in most cases trypsin, cleaves the peptide bonds, the heavy isotope is incorporated so all tryptic peptides will be labelled. The subsequent data analysis is identical to SILAC methods. A disadvantage of this approach is the relative expense of $H_2{}^{18}O$.

1.7.9.5 Dimethyl isotopic labelling

A further method uses standard and deuterium isotopes of formaldehyde to label the amino-terminus of peptides or the amino group of Lysine residues[428]. Again the isotopes are resolved by their mass differences allowing peptide level quantitation from the MS1 scan. A further limitation common to SILAC, ^{18}O and Dimethyl labelling is that a maximum of three samples can be compared per mass spectrometry analysis.

1.7.9.6 Isobaric peptide labelling

Isobaric peptide labels offer greater multiplexing capabilities with 4-plex or 8-plex (Isobaric Tag for Relative and Absolute Quantification; iTRAQ)[429] or 6-plex (Tandem Mass Tags; TMT)[430] commercial kits available. Both of these kits use the same underlying principle.

Each label consists of an amine reactive, N-hydroxy succinimide ester, balancing carbonyl linker and reporter ion components (Figure 1-7). Tryptic peptides form amide linkages with the labels via N-termini or lysine residues. A label with a different reporter is used for each different sample and all the samples are mixed prior to fractionation and LC-MS/MS. Each label has the same total mass and chromatographic properties and therefore the LC retention time and mass/charge (m/z) separation of each sample are not differentially affected during the MS1 scan[429].

Precursor ions are then sampled for MS/MS analysis and the ionized labeled peptides are fragmented with dissociation of the reporter ions from the balancing carbonyl linker. The peptide fragments are detected generating mass spectra in the usual manner. The reporter ions are also detected and as peaks at a predefined m/z. For a four-plex iTRAQ experiment the reporters are detected at 114.1, 115.1, 116.1 and 117.1 m/z[429]. For a sixplex TMT experiment, the reporters are detected at 126.1, 127.1, 128.1, 129.1, 130.1 and 131.1 m/z[430].

Assuming complete peptide labeling of each sample, the more abundant peptides within each sample will have accumulated more label. When equal amounts of each sample are mixed and subjected to LC-MS/MS together, those samples with a greater original concentration of a particular peptide will produce higher reporter ion peak intensities in the MS/MS scan. By comparing the relative reporter ion intensities, the relative peptide and therefore protein abundances in the original samples can be determined[431].

The multiplexing capabilities of isobaric labels are directly offset by the consequent dilution of each sample leading to challenges in identifying low abundance peptides[406]. Samples also have to be lysed and digested separately which has the potential to introduce error. In contrast, cell populations can be mixed prior to lysis in SILAC experiments. The quantitative accuracy and dynamic range offered by isobaric labels are excellent, however, surpassing SILAC in a direct comparison[417].

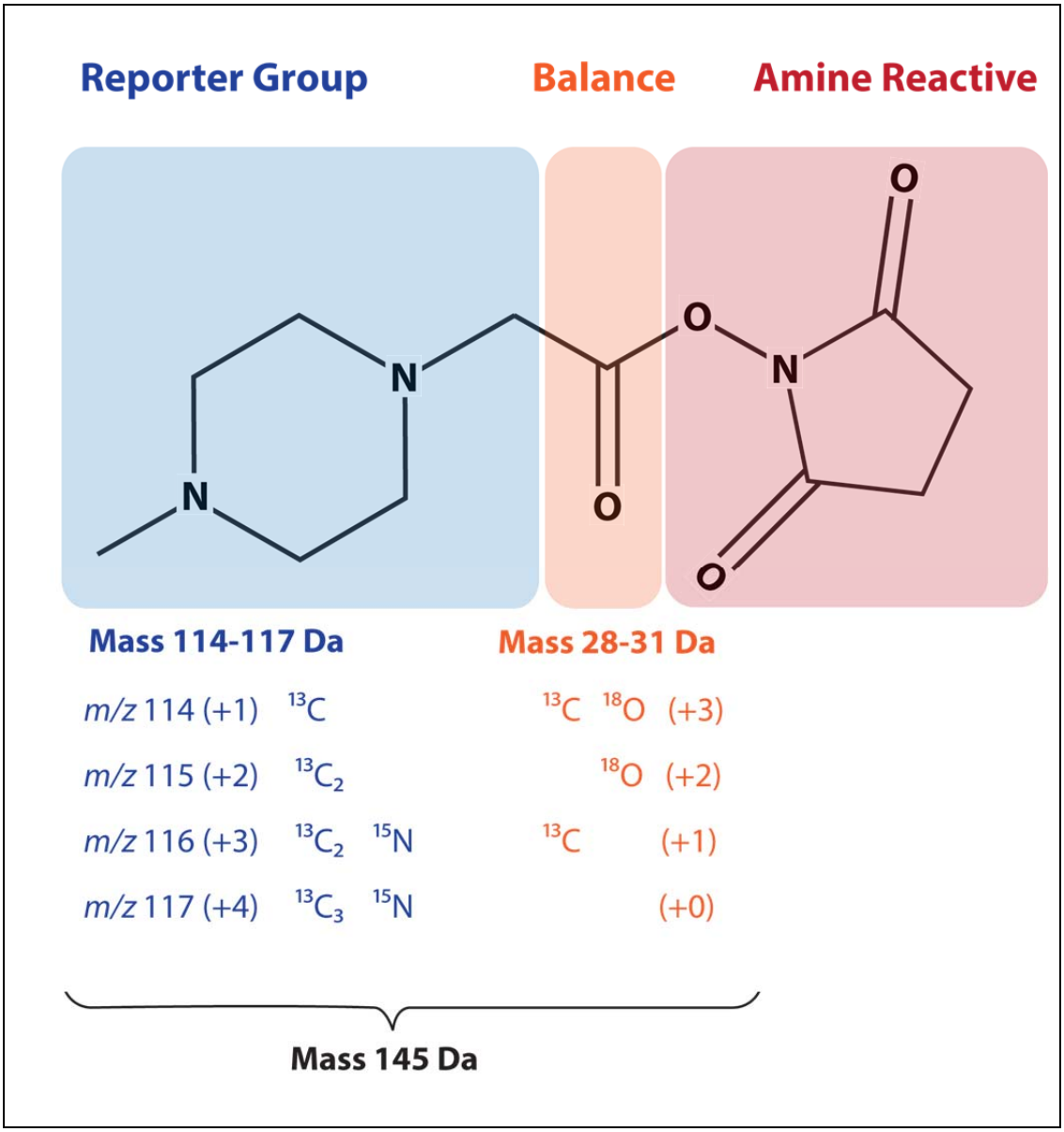

Figure 1-7 Schematic of the four-plex iTRAQ peptide label. A reporter group with a defined mass between 114 and 117 Da is connected to a balancing linker. Together the reporter and linkers have a fixed mass of 145 Da and they are connected to an amine reactive group which binds peptide amino-termini and lysine residues. The label is cleaved at the balancing linker during MS2 fragment ion generation to allow reporter ion detection. Figure adapted from reference[429]

1.7.10 Label-free Quantitative Shotgun Proteomics

All sample manipulation steps during a proteomic workflow reduce the data yield due to loss of proteins[432] and are additional sources of variation[433]. Eliminating the sample processing steps to incorporate labels for quantitation is clearly an advantage and underlies the rationale to develop label-free methods of quantitation.

The total number of spectra matched to each peptide contributing to a protein identification, termed the spectral count, has been reported to correlate with absolute protein abundance[434]. Various methods have been proposed to refine the spectral count such as normalising for protein length[435], or combination scores including peptide count and fragment-ion intensity[436]. For complex protein mixtures, spectral counts are still subject to significant between-run variability and are highly dependent on LC conditions and precursor ion selection. As a result the quantitative reproducibility of spectral counting is inferior to isobaric labelling methods[437].

An alternative relies on the capture of precursor ion intensity as a function of time to produce an ion chromatogram. The area under the ion chromatogram curve is linearly proportional to the peptide concentration[438]. Challenges exist in applying this method across LC-MS/MS runs to allow differential analysis as the same peptide ion must be identified and quantified despite background noise, co-eluting peptides causing signal overlap, technical variations in retention time and total protein loading among other factors[439].

A simultaneous advantage and disadvantage of both label-free approaches is the requirement to analyse one sample per LC-MS/MS run. This prevents sample dilution, provides maximum potential coverage and prevents the potential failure to identify dysregulated low abundance proteins that occurs with multiplexed approaches. By comparing conditions across separate LC-MS/MS runs, however, the inherent changes in LC performance and the stochastic nature of protein identification by shotgun proteomics both contribute to data heterogeneity. Until these concerns are addressed, labelling strategies will still be widely employed.

1.7.11 Considerations for Cancer Tissues Proteomics.

Proteomic methods have been most widely applied in human disease to the discovery of cancer biomarkers. A typical experimental design compares the proteome of multiple tissue biopsies derived from cancer and corresponding normal tissue. One challenge is to measure the proteome of samples that accurately reflect the *in vivo* expression profile. Tissue biopsies undergo hypoxic stress which leads to apoptosis and proteolysis with increasing time at room temperature[385]. For this reason,

biopsies are often snap-frozen in liquid nitrogen to temporarily inactivate enzymes prior to tissue lysis. A recently described alternative is the rapid heat inactivation of tissue biopsies[440]. For both methods, time is considered critical to prevent proteome degradation.

The heterogeneity of cancer biopsies has prompted the search for methods to isolate homogeneous cell populations from tissue samples. One such method, laser-capture microdissection (LCM), allows homogeneous cellular populations to be isolated from thin (<10 μm) frozen tissue sections[441]. Although allowing the discrete analysis of the complex cellular components that comprise the tumour microenvironment, the limited protein mass produced during microdissection necessitate multiple dissections to generate the 5,000-50,000 cells required for a typical shotgun proteomic experiment[442]. LCM therefore remains a labour intensive method that has yet to become widely adopted for whole proteome analysis.

A further challenge of studying the proteome of tissue biopsies is protein extraction. Membrane proteins are relatively hydrophobic and often prove insoluble in aqueous buffers. The use of detergents can enhance membrane protein recovery but are incompatible with mass spectrometry due to signal distortion. A hybrid method is reported to allow complete tissue proteome solubilisation using a detergent-containing buffer followed by buffer exchange into a mass-spectrometry compatible solution on molecular weight cut-off filters[443]. This method, termed filter-aided sample preparation (FASP) has also been successfully applied to LCM tissues[444]. If validated across multiple human tissue types this method could allow standardization between studies, a significant limitation in the current proteomics literature.

1.7.12 Quantitative Proteomics in OAC

Although quantitative proteomic methods have now been applied across many cancer types no studies have yet identified and quantified proteins from OAC tissue *de novo* with subsequent selection of dysregulated candidates (Table 1-3). Two studies have employed 2D gel electrophoresis of fresh frozen biopsy lysates. One compared protein spot intensities between OAC and matched normal mucosal biopsies[445] and the second compared spot intensities from tumours that did or did not show a

pathological response to chemoradiotherapy[446]. A further study has used a similar approach with protein separation and quantitation in the liquid phase and selective identification of dysregulated candidates[447]. Three studies have used MALDI-IMS to identify differential spectra features between tissue types and then gone on to subsequently identify a small number of these spectra.

The minimal protein expression data provided in these reports limits the comparisons that can be made across other studies or other cancer types. An additional advantage of generating proteomic expression data is that deranged signaling pathways can be identified using network analysis methods, potentially uncovering tractable therapeutic targets. There is therefore a need for larger scale quantitative proteomic studies in OAC.

Table 1-3 Proteomic Studies in Oesophageal Adenocarcinoma

Study Author	Patients	Method	Tissue Preparation	Squamous Samples	Gastric Samples	Barrett's Samples	ACC Samples	SCC Samples	Total Proteins Identified	Number of Proteins Dysregulated
Zhao[448]	6	LC-ESI TOF MS, Targeted LC-MS/MS	Fresh frozen biopsies	-	-	6	6	-	-	38 proteins
Yoo[449]	1	LC-ESI TOF MS, Targeted MALDI-MS	Fresh frozen biopsies	-	-	-	1	-	22 proteins	-
Peng[445]	8	2D Gels, Targeted MALDI-MS	Fresh frozen biopsies	2	2	-	8	-	-	23 dysregulated gel spots – 22 proteins identified.
Langer[446]	20	2D Gels, Targeted MALDI- and LC-MS/MS	Fresh frozen biopsies				20		-	Data for 4 proteins presented
Quaas[450]	477	MALDI-MSI	FFPE slides from a tissue microarray	-	-	-	300	177	72 different spectral features and 13 peptides identified.	-
Aichler[451]	23	MALDI-MSI. Targeted LC-MS/MS	Fresh frozen biopsies	-	-	-	23		-	22 spectral features, 6 proteins.
Elsner[452]	38	MALDI-MSI. Targeted LC-MS/MS	Fresh frozen biopsies	-	-	11	33			61 spectral features, 6 proteins
Streitz[453]	4	MALDI-MS	LCM of fresh frozen biopsies	-	-	4	4		8 spectral features	

1.8 Network Analysis

A key limitation of proteomic methods is the lack of complete proteome coverage[423]. Only a subset of the expressed proteome can be measured in any experiment and a further subset of this quantified. Biochemical properties and abundance play a small role in determining the likelihood of identification but otherwise this remains a stochastic phenomenon[454]. Identifying the changes in signalling underlying a phenotype is hampered by incomplete coverage of network or pathway components and difficulties in determining serial changes in protein expression[455].

Network analysis methods can address this by uncovering dysregulated pathways from incomplete expression data. This can be achieved if the underlying connectivity (edges) between genes or proteins (nodes) is described for that organism. The control of the majority of human cellular signalling pathways, however, remains under investigation and new models of regulation are regularly being uncovered[456]. Despite this, functional linkages have been described on a genome-wide scale for several organisms including humans. Functional linkages can refer to direct protein-protein interaction[457], co-regulation by a transcription factor[458], similar biological activity[459] or structural domain homology among other properties. In one example of a human functional linkage network, nodes are connected by functional linkages which summarise the evidence for a functional connection comprising protein-interaction data, co-expression, domain homology, shared ontologies and closely approximated genomic loci[460].

One approach to the use of functional linkage networks for proteomic data analysis is to generate clusters within the network using an algorithm that identifies highly connected nodes[461]. Due to the strategy of network construction, these clusters represent functionally connected genes. Dysregulated proteins identified by proteomics analysis are mapped onto these clusters along with a random resample from all identified proteins. Statistically enriched clusters can then be identified using the dysregulated protein list[462]. These clusters of functionally related genes represent a starting point for hypothesis directed protein expression mapping and subsequent confirmatory experiments.

1.9 Aims of This Project

Oesophageal adenocarcinoma is an increasing problem in the UK and current systemic therapies are ineffective for the majority of patients. The central aim of this study is to identify novel therapeutic targets for this cancer. A variety of strategies will be employed. The finding of a high frequency of TP53 mutation in OAC suggests a synthetic lethal approach may be feasible by targeting PLK-1.

Aim (i) ***To determine if PLK-1 represents a tractable therapeutic target in OAC using a panel of oesophageal cell lines representing the stages of oesophageal carcinogenesis.***

A key requirement for chemotherapeutic efficacy is to establish cancer cell cytotoxicity while minimising toxicity to non-transformed cells. Identifying proteins that are specifically expressed on the surface of OAC cells may allow effective immunotherapeutic approaches and enhanced clinical staging protocols. Proteomic strategies offer the potential to identify these proteins but mandate high quality tissue is used.

Aim (ii) ***To establish a human oesophageal tissue bioresource to enable therapeutic target discovery.***

Aim (iii) ***To establish the optimal methods for quantitative oesophageal tissue proteomics.***

Aim (iv) ***Using these methods – define cancer specific targets for development using an immunotherapeutic approach.***

Incomplete proteome coverage can limit the identification of dysregulated signalling pathways. Network analysis may complement quantitative proteomic strategies to reveal druggable targets.

Aim (v) ***Use network analysis tools to define potential targetable pathways in OAC.***

Chapter 2: Materials and Methods

2.1.1 Chemicals and Solutions

All stock solutions and additives used with cultured cells were purchased as sterile tissue-culture grade or were filtered using a 0.2 2µm syringe filter (Millex-GS, 0.22 µm, Millipore) and hand-held syringe. All chemicals and solutions were obtained from Sigma-Aldrich, Gillingham, UK unless otherwise stated.

2.1.2 Replicates

The number of experimental replicates undertaken has been indicated in the accompanying figure legends. Technical replicates were defined as a repeat of the experiment using material derived from the same tissue culture plate or biological material. Biological replicates refer to a repeat of the experiment using cells from a different tissue culture plate or biological material from a different patient.

2.2 Cell Lines

2.2.1 Cell Line Panels

Cell line based research is commonly criticised for a lack of relevance to tumours *in vivo*. This criticism is often based on the potentially confounding effect of artefacts induced by 2D culture. Despite these concerns, a panel of breast cancer cell lines has been shown to model the majority of genetic features of human breast cancer[89]. In addition, a panel of gastric cancer cell lines has revealed intrinsic molecular subtypes with differing drug sensitivities in human gastric cancer[90]. Tissue-specific cell line panels therefore have the capacity to provide disease-relevant model systems and yield biological insight.

2.2.2 Cell Line Sources

The OE19 and OE33 cell lines (passage number 24 and 13) were purchased from the European Collection of Cell Cultures[463]. The telomerase-immortalised (hTERT), non-transformed, non-dysplastic Barrett's epithelial cell line - CP-A[91], the OAC cell lines; FLO-1, JH-Eso-Ad1 and OACM5.1 C, and the Oesophageal Squamous Cell

Carcinoma cell line, KYSE-30, were generous gifts from Dr. Rebecca Fitzgerald (Cambridge). Primary oesophageal squamous cells; HEEpiCs were purchased from ScienCell research Laboratories[464]. The colorectal adenocarcinoma cell line, HCT116 was a generous gift from Dr. Bert Vogelstein (Baltimore) and the metastatic breast cancer cell line MCF-7 was purchased from the European Collection of Cell Cultures. All frozen cell stocks were maintained in liquid nitrogen until resuscitation by rapid thawing and dispersion into warmed complete medium. The characteristics of the oesophageal cells used are presented in Table 2-1.

2.2.3 Cell Line Propagation

All cell lines were maintained in a humidified incubator (Hera) at 37°C and 5% CO_2 and, with the exception of HEEpiCs, were grown on standard plastic tissue culture plates. The cell lines OE19, OE33, OACM5.1 C, and KYSE-30 were grown in RPMI 1640 medium (Invitrogen), FLO-1 and MCF7 cells were propagated in DMEM (Invitrogen) and HCT116 cells were grown in McCoys 5A medium (Invitrogen). RPMI, DMEM and McCoys 5A were supplemented with 10% (v/v) Foetal Calf Serum (FCS, Labtech International, Biosera) and 0.5 units/ml Penicillin + 500 ng/ml Streptomycin (P/S, Invitrogen) to make complete growth medium. The JH-Eso-Ad1 line was grown in MEM + 20% FCS and P/S. CP-A cells were maintained in Keratinocyte Serum Free Media (KSFM, Invitrogen) supplemented with recombinant epidermal growth factor (EGF, 5 µg/L), bovine pituitary extract (BPE, 50 mg/L) (both Invitrogen) and P/S.

For the propagation of HEEpiCs, tissue culture plates were first incubated with poly-L-lysine, diluted from a sterile stock of 10 mg/ml (ScienCell) to 2 µg/cm^2 in autoclaved water, for at least 1 hour at 37°C followed by aspiration of the poly-L-lysine solution. Cells were then seeded from rapidly thawn, frozen stocks into poly-L-lysine coated flasks containing warmed Epithelial Cell Medium-2 (EpiCM-2, ScienCell) supplemented with 1% Epithelial cell growth supplement-2 (EpiCGS-2, ScienCell) and P/S.

Table 2-1 Characteristics of Oesophageal Cell Lines and Primary Cells Used

Characteristics	Primary Cells	Cell line						
	HEEpiC	CP-A	OE19	OE33	FLO-1	OACM5.1 C	JH-Eso-Ad1	KYSE-30
Year Established (reference)	-	2003[91]	1993[465]	1993[465]	1991[85]	1996[466]	1997[467]	1992[468]
Derived From	Oesophageal Epithelium	Oesophageal Epithelium	Primary Tumour	Primary Tumour	Primary Tumour	Lymph Node Metastasis	Primary Tumour	Xenograft
Primary Site and Histology	Normal Squamous Oesophagus	Non-Dysplastic Barrett's	Gastric Cardia ACC	Lower Third Oesophageal ACC	Lower Third Oesophageal ACC	Lower Third Oesophageal ACC	Lower Third Oesophageal ACC	Middle Third Oesophagea l SCC
Neoadjuvant Treatment	-	-	No	No	No	No	No	No
Immortalised	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Generates Xenograft	No	No	Yes[85]	Yes[85]	Yes[469,470]	Yes[85]	Yes[85]	Yes[468]
TP53 Mutation	Wildtype	Wildtype	N310K Frame	C135Y	C277F	Frame	G266E	Exon 7 skipped[471]

Cell lines and primary cells were monitored daily using a bright-field inverted phase-contrast microscope and the medium changed every 48 hours or as required. Cells were maintained in an incubator specified for confirmed mycoplasma-free cells. Cell cultures were confirmed to be mycoplasma negative by periodic testing using the luciferase-based MycoAlert™ Mycoplasma detection kit (Lonza) as per the manufacturer's instructions.

2.2.4 Cell Subculture

For routine subculture, cell lines and primary cells were grown until 80-90% confluence. For OAC, MCF7 and HCT116 cells growing on 10 cm plates, the culture medium was aspirated and cells washed once with 10 ml of $Ca^{2+}$and Mg^{2+} free, sterile phosphate-buffered saline, (PBS, pH 7.4) followed by incubation for 5 minutes at 37^{o}C with 2.5 ml of 0.05% trypsin + 0.02% ethylene-diaminetetraacetic acid (EDTA). Once the cells had detached, 7.5 ml of complete medium was added and a single cell suspension was achieved by mixing before diluting 1:10 in fresh medium on a new tissue culture plate.

For subculture of CP-A cells, the growth medium was aspirated and cells washed once with 10 ml of PBS followed by incubation for 10 minutes at 37^{o}C with 2.5 ml of 0.25% trypsin, 0.02% EDTA. Once the cells detached, 7.5 ml of PBS + Soyabean Trypsin Inhibitor (STI, 250 mg/L, Invitrogen) were added and mixed thoroughly by pipetting. The cell suspension was centrifuged for 5 minutes at 1,000 rpm in a swing-bucket centrifuge and the supernatant discarded. The cell pellet was re-suspended in complete medium and dispersed onto a new tissue culture plate at a 1:10 dilution.

For subculture of HEEpiCs, the growth medium was aspirated and the cells washed with 10 ml of sterile PBS. A further 5 ml of sterile PBS was added to the cells followed by 5 ml of 0.05% trypsin, 0.02% EDTA and the cells incubated for 3 minutes at 37^{o}C. The trypsin solution was then aspirated and added to 5 ml of FCS. After 60 seconds incubation at 37^{o}C, 5 ml of trypsin neutralising solution (TNS, ScienCell) was added to the plate, mixed by pipetting, aspirated and then added to the FCS - trypsin mixture. This was repeated with a further 5 ml of TNS and the

plate reviewed by bright-field microscopy to ensure the majority of the cells had been harvested. The FCS – trypsin -TNS mixture was then centrifuged for 5 minutes at 1,000rpm, the supernatant discarded and the cell pellet resuspended in 10 ml of complete medium. The cell suspension was then split 1:5 to 1:10 and seeded into fresh poly-L-Lysine coated plates, prepared in advance, ensuring a minimum seeding density of 2,000 cells/cm^2. HEEpiC cells senesce after 12-15 passages and were therefore only used up until the 10^{th} passage.

For subculture or harvest of cell lines or primary cells growing on smaller plates, proportionally smaller volumes were used.

2.2.5 Cryopreservation

For long-term preservation of cell stocks, cells were maintained in a cryopreservation mixture and stored in liquid nitrogen. For OAC, MCF7 and HCT116 cells, a confluent 10 cm plate was trypsinised as previously but then centrifuged for 5 minutes at 1,000 rpm. The supernatant was discarded and the cells resuspended in 2.5 ml of complete medium. Cryovials containing a mixture of 0.5 ml of this cell suspension and 0.5 ml of cryopreservation solution (90% FCS: 10% DMSO) were then frozen overnight in an isopropanol bath at -80^{o}C followed by transfer to liquid nitrogen.

For CP-A cells and HEEpiCs the same procedure was followed for sub-culturing cells but the cell pellet was resuspended in 2.5 ml of complete medium. Cryovials were then prepared with 0.5 ml of cell suspension and 0.5 ml of cryopreservation mixture consisting of either 40% complete KSFM, 10% DMSO, 50% FCS (CP-A) or Defined Animal Component-Free Freezing Medium (HEEpiCs, TCS Cellworks) and were then frozen in the same fashion as OAC cells.

2.2.6 Harvesting of Cells

OAC cells were harvested for down-stream assays by washing once with ice-cold PBS followed by scraping into 500 µL of ice-cold PBS using a disposable rubber plate scraper and centrifugation of the cell suspension in a bench-top centrifuge at 3,000 rpm for 3 minutes. After discarding the supernatant, the harvested cell pellet

was either used immediately or frozen in liquid nitrogen. CP-A cells and HEEpiCs were harvested by trypsinisation as for subculture but cell pellets were resuspended in 0.5 ml of ice-cold PBS followed by centrifugation at 3,000 rpm for 3 minutes and subsequent identical treatment as OAC cell lines.

2.2.7 Cell Line Verification

Cell line misidentification and contamination was first reported over 4 decades ago[472]. Despite repeated warnings in the literature, this continues to be a problem[473]. Indeed, three lines widely used as oesophageal adenocarcinoma cells; SEG-1, BIC-1 and SK-GT-5, were recently identified to have been contaminated at an early passage and actually represent the lung cancer cell line H460, the colorectal adenocarcinoma cell line SW620 and the gastric fundus adenocarcinoma cell line SK-GT-2 respectively[85]. Unfortunately these cell lines continue to be inappropriately used as models for oesophageal adenocarcinoma with data from the discredited SK-GT-5 cell line recently published supporting oesophageal cancer-specific hypotheses[469].

Historical challenges in verifying the origin of a cell line have now been minimised with the use of short tandem repeat (STR) genotyping and availability of large STR genotype databases and multiplex PCR assays[474,475]. This method relies on heterogeneity between individuals in the number of tandem repeats at several genomic loci. By using high sensitivity capillary electrophoresis to quantify PCR-amplicon length from each locus, a "fingerprint" of alleles can be generated for a human DNA sample[476]. The combination of alleles from 10 loci produces a signature with a random match probability of less than 1 in a billion[477].

2.2.8 Confirmation of Cell Line Identify

For each oesophageal cell type, a confluent 10 cm plate or T75 flask was trypsinised, washed once in PBS and the cells pelleted. Genomic DNA was extracted from the pellet using the Wizard® SV Genomic DNA Purification kit (Promega) as per the manufacturer's instructions. Briefly, cell pellets were lysed at room temperature (RT) using the Wizard SV lysis buffer and lysates added to DNA purification spin columns and centrifuged at 13,000 rpm for 3 minutes. Spin columns were washed four times using the column wash buffer and then genomic DNA was eluted using

nuclease-free water warmed to 65°C. The DNA concentration of eluates was determined by measuring the absorbance at 260 nm (A_{260}) relative to absorbance at 280nm (A_{280}) using a NanoDrop 2000 UV-Vis spectrophotometer (ThermoScientific) and following the assumption that purified genomic DNA at a concentration of 50ng/µl has an absorbance of 1.0 absorbance units at 260 nm[478].

Genotypes were determined by short tandem repeat (STR) profiling using a commercial multiplex PCR assay covering 9 genomic loci; D7S820, D13S317, D16S539, D21S11, TH01, TPOX, vWA, CSF1PO and the Amelogenin locus for gender identification (Cell ID, Promega). Multiplex PCR was carried out as per manufacturer's instructions using a PTC-225 Peltier Thermal Cycler (MJ Research). DNA samples were de-identified by numbering and 2 ng of DNA used as a template in a final reaction volume of 25 µL. A positive control sample of K562 DNA and a no template control were included as separate samples. PCR products from each sample were then resolved by capillary electrophoresis (3130xl, Applied Biosystems) after spectral calibration using the PowerPlex Matrix Standards (Promega) and mixing with an internal lane standard. Detected amplicon peaks were manually binned as alleles using the custom bin reference files (downloaded from www.promega.com). For each sample, all the alleles confidently identified for each locus were recorded and then searched against the STR genotype database and published oesophageal STR genotypes[85,474]. The origin of each sample was then predicted, blind to the original sample origin. Samples were then re-identified and predictions and sample origins compared.

2.2.9 Cell Counting

To generate approximately consistent starting cell numbers for *in vitro* experiments across cell lines and conditions, cells were seeded at defined densities. Cells were trypsinised from adherent cultures and re-suspended in 10 ml of complete medium. A 20 µL aliquot of this suspension was diluted 1:1 with Trypan Blue and incubated for 3 minutes at RT. Cells were then counted using a haemocytometer with the number of viable cells per ml defined as; 2 x 10^4 x (the number of non-blue staining cells counted per 4x4 haemocytometer grid). The mean of at least 4 counts was

taken and the original cell suspension diluted down to the desired final concentration with complete medium.

2.2.10 X-Ray and Ultraviolet light (UV) exposure

Cells were trypsinised and seeded in 6 cm plates at an initial density of 500,000 cells/plate. This initial density ensured plates were sub-confluent at the time of harvest. Growth medium was changed 24 hours after seeding and experimental treatments were performed once plates were ~80% confluent (approximately the 4th day after seeding). A sealed cabinet x-ray system (Faxitron) was used to deliver a calibrated gamma radiation dose at a rate of 2 Gy/min. For UV-C treatment, cells were irradiated in a closed cabinet system at 70-80% confluence in 6 cm plates with the lids removed. Cells were then harvested at intervals up to 24 hours after treatment by scraping into ice-cold PBS and centrifugation at 3,000 rpm for 3 minutes followed by lysis of the cell pellet for protein extraction.

2.2.11 Positive Control for p53 Response to DNA damage

As a control for the effect of gamma irradiation on p53, a sub-confluent 10 cm plate of HCT116 cells was treated with gamma irradiation (10 Gy). Cells were harvested by trypsinisation at 24 hours, pelleted and lysed using urea buffer.

2.3 Protein Assays

2.3.1 Cell Pellet Lysis

Cell pellets were lysed in urea buffer comprising; 6.24 M Urea, 0.1 M Dithiothretiol (DTT), 0.05% Triton X-100, 25 mM NaCl, 20 mM HEPES-KOH (pH7.6), Complete Protease Inhibitor Cocktail (Roche), 5 mM NaF, 2 mM Na_3VO_4, 2.5 mM $Na_4P_2O_7$ and incubated on ice for 30 minutes. Lysates were then clarified by centrifugation at 13,000 rpm for 10 minutes and supernatants reserved and diluted to 2 µg/µL in lysis buffer after protein concentration determination.

2.3.2 Protein Concentration Determination

A Bradford assay was used to determine the protein concentration of cell lysates[479]. A dilution series of concentration standards across the range 4 mg/ml to 0.25 mg/ml was prepared by dissolving bovine serum albumin (BSA) in the lysis buffer used for

protein extraction. Bradford reagent was prepared as a stock (0.1 g/L Coomassie Blue G-250 in 8.5% Phosphoric Acid and 5% Ethanol) and stored protected from light at 4°C[479]. Bradford reagent was mixed with standards, samples of unknown concentration and lysis-buffer alone at a ratio of 200 µL: 1 µL, aliquoted in at least technical duplicate into the wells of a clear 96-well plate and incubated for 5 minutes at room temperature (RT) with vigorous shaking. Sample absorbance at 595 nm (A_{595}) was subsequently measured using a VICTOR^3 1420 multi-label plate reader (Perkin-Elmer). Unknown sample protein concentrations were inferred by interpolation on a linear regression line created from the BSA standard absorbances after subtraction of background absorbance from lysis-buffer only mixtures. Cell lysates were then diluted to 1.25 µg/µl or 2.5 µg/µl in distilled water (dH_2O).

2.3.3 SDS-PAGE

Cell lysates were further diluted by the addition of 5x Sample Loading Buffer (SLB, Table 2-2) at a ratio of 4 µL lysate: 1 µL SLB and heated at 95°C for 5 minutes and then cooled on ice. Recombinant purified p21 and MDM2 protein stocks were generous gifts from Dr. Jennifer Fraser (Edinburgh Cancer , Research Centre, Edinburgh) and Dr. Erin Worrall (Edinburgh Cancer Research Centre, Edinburgh) respectively. Purified proteins were dissolved in 1xSLB and maintained at -80°C until required when 1 µL, diluted 1:10 in 1xSLB, was used.

Lysates were loaded into a stacking gel and resolved by denaturing polyacrylamide gel electrophoresis (PAGE) using Bio-Rad gel apparatus and power packs, 1x sodium dodecyl sulphate (SDS) running buffer (Table 2-2) and a constant voltage of 180 V into polyacrylamide running gels (Table 2-2) at the desired acrylamide concentration (8-15% v/v). Electrophoresis was continued until the first protein molecular weight standard marker (Fermentas) reached the bottom of the gel. Gels were then incubated with Coomassie stain for 60 minutes at RT and then overnight in destain solution, rinsed in dH_2O, and dehydrated using a vacuum gel dryer (SCIE-PLAS Gel Dryer, GD 4534, Scientific Laboratory Supplier) or used for western blotting.

Table 2-2 Buffer and Solution Constituents

5 x Sample Loading Buffer	1x Running Buffer
250 mM Tris HCl (pH 6.8)	25 mM Tris HCl
50% v/v Glycerol	0.2 mM Glycine
5% w/v SDS	0.1% w/v SDS
0.5% w/v Bromophenol Blue	in dH_2O
250 mM DTT	
in dH_2O	1x Transfer Buffer
	25 mM Tris HCl
Running Gel	0.2 mM Glycine
30% Acrylamide (Protogel)	20% v/v Methanol
(Diluted to required percentage v/v)	in dH_2O
375 mM Tris HCl (pH 8.8)	
0.1% w/v SDS	3% BSA in TBS-T
0.1% w/v APS	3% w/v BSA
0.04% v/v TEMED	10 mM β-Glycerophosphate
in dH_2O	in Tris -buffered saline with 0.1% v/v Tween-20
Stacking Gel	5% Milk in PBS-T
5% v/v Acrylamide mix	5% w/v non-fat dry milk (Marvel)
125 mM Tris HCl (pH 6.8)	10 mM β-Glycerophosphate
0.1% w/v SDS	in Phosphate-buffered saline with 0.1% v/v Tween-20
0.1% w/v Ammonium Persulphate (APS)	
0.04% v/v TEMED	Coomassie Stain
in dH_2O	0.1% Coomassie Blue G-250
	10% v/v Glacial Acetic Acid
Destain Solution	40% v/v Methanol
20% v/v Methanol	in dH_2O
10% v/v Glacial Acetic Acid in dH_2O	

Abbreviations – TBS-T (Tris-buffered Saline + 0.1% (v/v) Tween-20), PBS-T (Phosphate-buffered Saline + 0.1% (v/v) Tween-20)

Lysates were loaded into a stacking gel and resolved by denaturing polyacrylamide gel electrophoresis (PAGE) using Bio-Rad gel apparatus and power packs, 1x sodium dodecyl sulphate (SDS) running buffer (Table 2-2) and a constant voltage of 180 V into polyacrylamide running gels (Table 2-2) at the desired acrylamide

concentration (8-15% v/v). Electrophoresis was continued until the first protein molecular weight standard marker (Fermentas) reached the bottom of the gel. Gels were then incubated with Coomassie stain for 60 minutes at RT and then overnight in destain solution, rinsed in dH_2O, and dehydrated using a vacuum gel dryer (SCIE-PLAS Gel Dryer, GD 4534, Scientific Laboratory Supplier) or used for western blotting.

2.3.4 Western Blotting

Resolved proteins and markers were transferred from gels to Hybond-C extra nitrocellulose membranes (Amersham Bioscience) in 1x transfer buffer using western blotting cassettes (Bio-Rad) and a constant current of 300 mA for 90 minutes. Membranes were then immediately blocked in 3% BSA in TBS-T or 5% Milk PBS-T for 90 minutes at RT with gentle shaking.

2.3.5 Primary Antibody Conditions

After blocking, membranes were incubated with primary antibodies as summarised in Table 2-3 and then washed four times, each for 5 minutes in 10 ml of PBS-T.

2.3.6 Secondary Antibody Conditions

For secondary detection of nitrocellulose membrane-bound primary antibodies, goat anti-mouse or goat anti-rabbit antibodies conjugated to either the IRDye™ 680RD (anti-mouse^{680} or anti-rabbit^{680}; excitation 679nm, emission 702nm) or IRDye™ 800CW (anti-mouse^{800} or anti-rabbit^{800}; excitation 778 nm, emission 806 nm) near-infrared fluorophores were used (Li-Cor Biosciences) at 1/2000 dilution in blocking buffer for 2 hours at RT protected from light. Blots were then washed four times, each for 5 minutes in PBS-T at RT protected from light and then imaged using a dual fluorophore near-infrared detection system (Odyssey SA, Li-Cor Biosciences).

2.3.7 Quantitative Western Blotting

Western blotting has traditionally relied on enhanced chemiluminescence (ECL)-based detection of bound secondary antibodies on either X-ray film or charge-coupled device (CCD) cameras and subsequent densimetry-based measurement for relative quantitation. This method has been considered semi-quantitative due to the

time-dependent signal variation as the ECL substrate is consumed, and the variability in band quantitation by densitometry[480]. In contrast, the detection of fluorescence from fluorophore-conjugated secondary antibodies offers improved quantitative accuracy and can provide a linear response over a 1,000-fold dynamic range[481]. The simultaneous use of both mouse and rabbit primary antibodies and secondaries with differing fluorophores allows relative quantitation of phosphorylated and non-phosphorlyated protein isoforms from the same blot or from proteins isoforms that are present in the same band on ECL western blots[482,483].

Cellular proteins and the membranes used for western blotting exhibit autofluorescence in the visible light range but this is significantly reduced in the near-infrared spectrum[484]. The use of fluorophore-conjugated secondary antibodies with emission spectra in the near-infrared range allows the detection of proteins by western blotting with similar or even improved sensitivity to enhanced chemiluminescence-based (ECL) methods and improved signal to noise ratio[480].

For quantitative western-blotting, membranes soaked in PBS-T were placed protein-side down in a glass membrane carrier (Li-Cor Biosciences) and bubbles between the membrane and glass extruded by gentle pressure. The membrane was then scanned on a Li-Cor Near Infrared Scanner (Odyssey, Li-Cor Biosciences) at 100 μm resolution as per manufacturer instructions with gain settings determined after a preview scan to ensure the detected signal was not saturated. Bands were manually selected from scanned images using the Odyssey SA software (Li-Cor Biosciences). Relative band intensities were measured using the integrated intensity method to determine intensity relative to band area. The same area was measured for all bands, ensuring in each case that the entire band was incorporated. No image manipulation was used prior to intensity calculation. The integrated intensity for β-Actin was measured on each blot as a loading control. For comparison of protein expression between conditions or cell lines, the protein expression relative to β-Actin was used.

For illustration of relative protein expression, the brightness and contrast of scan images were adjusted to allow visualization of both low and high abundance bands without intensity saturation and images exported in tagged image file format (TIFF).

The protein molecular weight markers were detectable as bands in the 700 nm channel and therefore provided a size reference for bands on scanned blots. Following image acquisition, membranes were either discarded or washed three times in PBS-T for 20 minutes each wash and then re-probed with further antibodies.

2.4 Cell Viability Measurement

2.4.1 CellTiter-Glo

CellTiter-Glo™ (Promega) is a luminescence assay which lyses cells and produces light in proportion to the ATP released. The very short half-life of free extracellular ATP allows viable cell number to be accurately measured with this method although this is intrinsically an endpoint assay[485]. Cells were seeded into 96 well clear-bottomed, black-walled plates (Costar) at variable densities with a final volume of 100 µL/well. A media-only control was included. At the required time point, plates were removed from the incubator and allowed to equilibrate to RT for 30 minutes. CellTiter-Glo® reagent (100 µL, Promega) was added to the wells and mixed on an orbital shaker for 2 minutes. The plates were left on the bench to equilibrate for a further 10 minutes before luminescence measurement using a Fluoroskan Ascent FL plate reader (Labsystems) with a 1000 ms integration time per well. Following luminescence acquisition, the contents of each well were transferred to a 96 well, white-walled, white-bottomed plate (Costar) and a further reading taken. The luminescence from media-only controls was used for background correction of the other readings.

2.4.2 AlamarBlue

AlamarBlue™ (Sigma), in contrast, can be used for serial measurements of cell number since the reagent can be washed-off without affecting viability[486]. This inert soluble resazurin salt is reduced to the fluorophore resorufin by the metabolic reductase capacity of viable cells. This reduction is proportional to the number of viable cells and can be measured as a change in fluorescence.

Cells were seeded into 96 well clear-bottomed, black-walled plates (Costar) at 2 - 10,000 cells/well with a final volume of 100 µL/well. A media-only control was

included to allow subsequent background correction. After 24 hours, 10 µL of sterile-filtered AlamarBlue (Invitrogen) was added and the plates incubated for a further 60 minutes at 37°C. Plates were imaged using a standard fluorescence intensity protocol (excitation 540 nm, light 5%, Gain 100, Flashes 10, emission recorded at 590 nm) on a Wallac Envision 2101 plate reader (Perkin Elmer). Following measurement, the medium and AlamarBlue reagent were aspirated and the wells washed once with fresh medium.

Drug aliquots (see 2.4.3 Drug Stocks) or vehicle were added to wells and, after 96 hours incubation, a further AlamarBlue measurement was taken using the above protocol. Fluorescence values were corrected for background fluorescence from medium only control wells. Final background-corrected fluorescence was then compared with initial background-corrected fluorescence on a well-by-well basis and the ratio of final/initial calculated and converted to a percentage of the initial value. Values >100 were taken to indicate cell proliferation and values <100 to indicate cytotoxicity. Trimmed mean percentages +/- standard errors were calculated and normalised to DMSO-treated cells. The performance of the AlamarBlue assay was assessed using the z-factor calculated as;

$$Z = 1 - \frac{3(\sigma_a - \sigma_b)}{(\bar{a} - \bar{b})}$$

Equation 1 Calculation of the Z-factor for assay performance (from reference [487]). σ= standard deviation, a = positive control, b = negative control.

Table 2-3 Antibody Conditions for Western Blotting

Antibody Directed Against	Source, Reference Number	Species of Origin	Clonality	Antibody Diluent	Incubation concentration	Incubation time, temperature	Predicted Band(s) Size
AGR2	Abnova, H00010551-MO3	Mouse	Monoclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	17kDa
AGR2	Atlas Antibodies, HPA007912	Rabbit	Polyclonal	5% Milk, PBS-T	1/1000	2 hours, RT	17kDa
AURKA	New England, #4718	Rabbit	Monoclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	47kDa
AURKA (cTerminus)	Abcam, ab52973	Rabbit	Monoclonal	5% BSA, TBS-T	1/8000	16 hours, 4°C	47kDa
β-Actin	Sigma, AC-15	Mouse	Monoclonal	5% Milk, PBS-T	1/5000	2 hours, RT	42kDa
BAX	Santa Cruz, N20	Rabbit	Polyclonal	5% Milk, PBS-T	1/500	48 hours, 4°C	23kDa
FBP1	Atlas Antibodies, HPA005857	Rabbit	Polyclonal	5% Milk, PBS-T	1/250	16 hours, 4°C	37kDa
GAPDH	Abcam, ab9484	Mouse	Monoclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	40kDa
GRP-78	Abcam, ab21685	Rabbit	Polyclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	78kDa
GSTO-1	Abcam, 129106	Rabbit	Monoclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	28kDa
HSP-27	Abcam, ab114067	Mouse	Monoclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	27kDa
MDM2	Moravian Biotech, 2A10	Mouse	Monoclonal	5% Milk, PBS-T	1/500	16 hours, 4°C	90kDa (MDM2) and ~170kDa (non-MDM2) band[488]
P21(WAF1)	Calbiochem, EA10	Mouse	Monoclonal	5% Milk, PBS-T	1/500	48 hours, 4°C	21kDa
P53	Moravian Biotech, DO-1	Mouse	Monoclonal	5% Milk, PBS-T	1/5000	2 hours, RT or 16 hours, 4°C	53kDa
P53	Moravian Biotech, CM1	Rabbit	Monoclonal	5% Milk, PBS-T	1/1000	16 hours, 4°C	53kDa
PLK-1	Abcam, ab17056	Mouse	Monoclonal	5% Milk, PBS-T	1/2000	16 hours, 4°C	67kDa
Rho GDI-2	Abcam, ab88317	Rabbit	Polyclonal	5% Milk, PBS-T	1/500	16 hours, 4°C	23kDa
SAMHD-1	Abcam, ab119751	Mouse	Monoclonal	5% Milk, PBS-T	1/250	16 hours, 4°C	72kDa
TGM-3	Sigma, HPA004728	Rabbit	Polyclonal	5% Milk, PBS-T	1/500	16 hours, 4°C	77kDa
Anti-mouse$^{680 \text{ and } 800}$	Licor Biosciences	Goat	Polyclonal	5% Milk, PBS-T	1/2000	90 minutes, RT	Mouse primary
Anti-rabbit$^{680 \text{ and } 800}$	Licor Biosciences	Goat	Polyclonal	5% Milk, PBS-T	1/2000	90 minutes, RT	Rabbit Primary

2.4.3 Drug Stocks

Table 2-4 Drug Stocks Used For In Vitro Experiments

Drug Name	Source	Stock Concentration (Solvent)
PLK-1A	Provided as purified dry compound after materials transfer agreement with Cyclacel Pharmaceuticals Incorporated, Dundee	10 mM (DMSO)
BI2536	Axon Medchem BV, Groningen, Netherlands	10 mM (DMSO)
Cisplatin	Sigma Aldrich, Gillingham, UK	5 mM (PBS)
Nutlin-3A	Alexis Biochemicals, Exeter, UK	43 mM (DMSO)

2.4.4 Comparison of Drug Effects

Best-fit curves were iterated using the relative viabilities across a drug dose range and a variable-slope, four-parameter, sigmoidal curve-fitting algorithm (GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego, California, USA). The IC_{50} and corresponding 95% confidence intervals were automatically generated using this algorithm, the best-fit curve and the R^2 value. The best-fit line was also used to determine the maximal drug effect relative to baseline, summarised as; reduction in proliferation, cytostasis or cytotoxicity.

2.4.5 Colony Formation Assay

Cells were seeded in 10 cm plates in complete medium, allowed to adhere for 24 hours and then the medium exchanged and supplemented with drug diluted from DMSO stock or DMSO only. Cells were harvested after 72 hours by trypsinisation and defined numbers of trypan blue excluding (viable) cells seeded in 6cm plates in complete medium only.

Plates were incubated in standard conditions with exchange of medium every 3-4 days until DMSO control plates approached confluence. The medium was then aspirated, plates washed once with 5 ml of PBS and cells fixed with 5 ml of ice-cold methanol for 5 minutes at RT. The methanol was aspirated and plates left to air-dry for 1 hour at RT.

To stain colonies, plates were incubated with Giemsa stain (0.05% w/v in buffered methanol solution, pH 6.8, Sigma) for 20 minutes at RT, gently washed in tap water and then left to air dry. Due to the diverse growth patterns of the oesophageal cells used, raw colony counts were considered uninformative and for comparison of drug effects, plates were simply photographed for visual comparison. All plates at the same seeding density were fixed and stained at the same time.

2.4.6 Death Commitment Assay

Baseline viable cell number was calculated using AlamarBlue, 24 hours after seeding into 96 well, poly-L-lysine coated plates (final volume 100 µL; OE33, CPA 2,000 cells/well; OE19 5,000 cells/well). Drug dilutions from DMSO stock or DMSO (1/10,000) were added and an interim AlamarBlue assay performed after a further 24 hours. After 1 wash to remove the AlamarBlue, fresh medium or medium supplemented with drug or DMSO was added to the wells and culture continued for a further 72 hours before a final Alamar assay was performed. Interim viability was normalised relative to baseline and final viable cell number calculated relative to the interim value so that when plotted, the x-axis represents no change from the interim value and the y-axis was scaled so a 100% change in viability was equivalent to the baseline cell number.

2.4.7 In-Cell Western

Cells were grown in poly-L-lysine coated 96 well plates. After completing an AlamarBlue assay, growth medium was aspirated, cells washed in 200µL PBS and then fixed with 100% methanol for 5 minutes at 4^{o}C. Cells were then permeabilised by gentle shaking in PBS and 1% Triton X-100 for 5 minutes with this step repeated 5 times.

Wells were blocked with 5% BSA and 0.1% (v/v) Tween-20 for 2 hours at RT. Primary antibody incubation were for 16 hours (hrs) at 4^{o}C in blocking buffer with antibodies to PLK-1 (ab17056, 1/200), the c-terminus of AURKA (ab52973, 1/1000), GAPDH (ab9484, 1/200) or full length AURKA (#4718, 1/200). Secondary incubations were with anti-mouse^{800} or anti-rabbit^{800} antibodies (Li-Cor, 1/2000, 2

hrs RT, protected from light) or the DNA stain, DRAQ5 (Li-Cor, 1/10,000, 2 hrs RT, protected from light).

Plates were scanned on the Li-Cor Near Infrared Scanner (Odyssey, Li-Cor Biosciences). The 800 nm integrated intensity values were background-corrected using the fluorescence readings from secondary antibody-only treated wells. The 700 nm DRAQ5 integrated intensity values were background corrected using readings from wells seeded with medium only (no cells). The ratio of background corrected 800nm/700nm intensity or 800nm /AlamarBlue intensity was calculated.

2.4.8 Real-time proliferation measurements

In 2D culture, the area covered by cells increases proportionately with cell number. This property is exploited in the xCELLigence system (Roche Diagnostics, Burgess Hill, UK) where plates lined with interdigitated gold electrodes are used to monitor impedance changes across the growth surface. Changes in impedance correlate directly with changes in cell number and allow real-time, label-free monitoring of proliferation in culture[489].

Real-time proliferation experiments were conducted on an xCELLigence RTCA DP instrument (Roche Diagnostics, Burgess Hill, UK) maintained in a humidified incubator at 37^{o}C, 5% CO_2. For each experiment, 100 µL of complete medium was added to each well of a sterile DP 16 well plate (E-plate, Roche Diagnostics) and sterile water added to the well surround to aid local humidification of the well. The plate was allowed to equilibrate to 37^{o}C on the RTCA reader before a baseline impedance measurement was taken.

Cells were seeded into wells with a final volume of 180 µL per well. An impedance reading was taken every 15 minutes until the end of the experiment and background corrected to derive a cell-index. Approximately 24 hrs after cell seeding, 20 µL of drug or DMSO was added and mixed by gentle pipetting. For wash-off experiments, cells were incubated in drug or DMSO for 40-50 hours and then the medium aspirated, cells gently washed once with 200 µL fresh medium before a final incubation in 200 µL fresh medium for a further 90 hours.

Experiments were terminated when growth curves in any of the wells plateaued or at 96 hours after drug wash-off, whichever was earlier. At the end of the experiment, wells were checked by light microscopy to ensure there was no obvious infection and that cells had not become confluent during the experiment. Cell index measurements were normalized to a point one to two hours after a change in conditions to exclude short-term fluctuations in impedance. Changes in cell-index were assumed to represent changes in cell number.

2.4.9 Validation of xCELLigence Viability Measurements

Cells (CP-A, KYSE-30; 2500 cells per well) were seeded in E-plates and allowed to adhere for 24 hours before drug or DMSO addition. The mean final cell index at 96 hours was calculated for each condition from technical duplicate wells. Plates were removed from the incubator and allowed to equilibrate to RT for 30 minutes. Medium was then aspirated and cells lysed using 100 µL of CellTiter-Glo reagent per well. After mixing on an orbital shaker for 2 minutes, lysates were transferred to a 96 well, white-opaque plate (Costar) along with a 1:1 mixture of CellTiter-Glo reagent and complete medium as a background control. Plates were incubated at RT for a further 10 minutes before luminescence was recorded as previously described (2.4.1).

2.4.10 Colony Outgrowth Assay

Cells (CP-A, OE33 - 2,500 cells per plate; OE19 - 10,000 cells per plate) were seeded in 6 cm plates and allowed to adhere for 24 hours. Drug dilutions or DMSO were added for 16 hours, plates washed once and then subsequently incubated in complete medium. Plates were incubated in standard conditions with exchanges of medium every 3-4 days until DMSO control plates approached confluence (12 days). Colonies were then fixed and stained as previously described (2.4.5).

2.5 Gene Expression

2.5.1 RNA extraction from cell lines

Bench surfaces and pipettes were cleaned with RNaseZap (Ambion) to inactivate environmental RNases and standard precautions taken to minimize contamination during RNA purification and preparation. RNA was extracted from cell pellets using the RNeasy Minikit (Qiagen) as per manufacturer's recommendations.

Briefly, pellets were lysed directly using 600 µL of Buffer RLT by pipetting and lysates centrifuged through Shredder mini spin-columns (Qiagen) to ensure complete lysis. The flow-through was mixed 1:1 with 70% (v/v) Ethanol in RNase-free H_20 before the solubilized RNA was bound to silica membranes by centrifuging over RNeasy mini-spin columns (Qiagen). After sequential spin-column washes, RNA was eluted into 30 µL of RNase-free H_20 and the RNA concentration determined by measuring the A_{260} on a NanoDrop 2000 UV-Vis spectrophotometer following the assumption that purified RNA at a concentration of 40 ng/µl has an absorbance of 1.0 absorbance units at 260 nm[478]. An A_{260}/A_{280} ratio of >1.8 was deemed acceptable RNA purity, otherwise extraction was repeated on a further cell pellet. Eluted RNA was snap frozen in liquid nitrogen and then stored at -80^{o}C until use.

2.5.2 Reverse Transcription of RNA to cDNA

Template complementary DNA (cDNA) was generated for each sample from 2 µg of total RNA in 7 µL of RNase-free H_2O. Sample RNA was heated to 65^{o}C for 5 minutes and cooled on ice before addition to a reaction mixture (final volume 20 µL) consisting of;

Sample RNA	100 ng/ µL
Reverse Transcriptase buffer (Qiagen)	1x
dNTP mixture (Qiagen)	0.5 mM for each dNTP (Qiagen)
oligo (dT) 12-18 primer (Invitrogen)	1 µM
Protector RNase inhibitor (Roche)	10 Units
DTT (Sigma)	10 mM
Omniscript Reverse Transcriptase (Qiagen)	4 Units

This mixture was incubated at 37°C for 1 hour to yield cDNA which was quantified using a Nanodrop as previously described and then stored at -20°C until required.

2.5.3 Quantitative real-time Polymerase Chain Reaction (qRT-PCR)

To accurately measure relative changes in mRNA levels between conditions and cell lines, qRT-PCR is often performed. Robust commercial assays allow the quantification of relative cDNA through the real-time measurement of the fluorescence derived from reporter dyes incorporated into gene-specific primers.

Solaris™ (ThermoScientific) assays utilise a quencher-primer-fluorophore combination where fluorescence is only detected as primers specifically anneal to the complimentary DNA sequence and the quencher dissociates from the fluorophore[490]. With progressive PCR cycles, the cDNA incorporating, and between, the primer annealing sites is amplified. The primers are in excess in the reaction and therefore the fluorescence measured is proportional to the cDNA concentration[491]. At a set cDNA concentration, the primer fluorescence passes the limit of detection by the qRT-PCR instrument and with each increase in template concentration the fluorescence increases proportionately[492].

At a certain point, the cDNA, primers, and nucleotides are at optimal concentrations to allow maximally efficient DNA polymerisation. At this point, termed the threshold cycle (C_t), there is an exponential increase in cDNA copy number and therefore fluorescence with each successive PCR cycle. The C_t is proportional to the starting cDNA concentration and can be used as a surrogate for RNA abundance.

The C_t is sensitive to variable cDNA loading prior to the PCR reaction. The gene expression relative to an invariant gene is therefore used; the ΔC_t. The metabolic gene GAPDH is often considered as a control gene, yet dynamic changes in GAPDH expression have been reported and it is therefore considered inappropriate as a qRT-PCR control[493]. Given the problems in selecting one invariant house-keeping gene for all reactions, the geometric mean of the C_t from two controls can be used[494].

In a 100% efficient reaction, each PCR cycle would double the cDNA concentration and therefore a difference of 1 C_t represents a two-fold difference in the RNA

amount. The RNA abundance relative to the baseline condition can therefore be determined as $2^{-\Delta\Delta Ct}$. With a lower efficiency this becomes $(1+efficiency)^{-\Delta\Delta Ct}$.

The PCR amplification efficiency can be determined by calculating the gradient of the linear regression line passing through the ΔCt values from a template dilution series subjected to qRT-PCR using the primer of interest. The amplification rate when the template concentrations are $\log_{10}$ transformed can then be calculated as[495];

$$Amplification\ rate = 10^{-1/gradient}$$

Where Amplification rate = Efficiency + 1.

Pre-designed gene expression assays were used for all genes (Solaris, ThermoScientific). Primer sequences are provided in (Table 2-5). To compare relative gene expression between conditions, the $\Delta\Delta C_t$ method was used[496]. Both β-Actin and Ribosomal protein components have shown stable expression levels between normal oesophageal and cancer tissue[497]. Ribosomal Protein S18 (RPS18) and β-Actin were therefore used to control for differences in cDNA loading between experiments.

Each cDNA sample was run in technical quadruplicate for each primer with 125 ng of cDNA used per reaction. A 2x mastermix was created for each primer containing the Solaris qRT-PCR primer, and a pre-mixed 2x solution of DNA Polymerase, dNTPs and reaction buffer (Solaris Mastermix, ThermoScientific). Primer mastermix and cDNA were mixed in a 1:1 ratio by pipetting in white PCR tubes (BIOplastics, Netherlands) maintained on ice and then sealed using clear thermostable caps (BIOplastics, Netherlands).

For each qRT-PCR reaction, a no template control was included for each primer by substituting RNase free H_2O for cDNA and an RNA only control was included for each sample by substituting 125 ng of total RNA (used as the input for the reverse transcriptase reaction) for cDNA. The primers for the genes (RPS18) and β-Actin were included as controls for all samples.

Reactions were commenced by heating to 95°C for 15 minutes (PTC-200, Thermal Cycler, MJ Research) followed by a PCR cycle consisting of; denaturation at 95°C for 15 seconds, combined annealing and extension at 60°C for 60 seconds with measurement of FAM fluorescence at the end of each annealing stage (Chromo4 Continuous Fluorescence Detector, MJ Research). A total of 40 PCR cycles were completed and then a melting curve obtained for the PCR products by stepwise increment of the reaction temperature from 60-95°C at 1°C per minute with fluorescence measurement after every temperature increment.

The melting curves were reviewed for all wells and only those with smooth Gaussian curves consistent with the other wells using that primer were included for subsequent data analysis. The no template control wells and RNA only wells were checked to ensure there was no fluorescence detected above background or melting curve consistent with PCR amplification.

Once the wells were selected for subsequent gene expression analysis, a fluorescence threshold was manually selected incorporating the exponential phase of fluorescence amplification in all wells. The cycle at which each sample crossed this threshold was automatically determined by the instrument software (Opticon 2, MJ Research) and was recorded as the Ct. The mean Ct, standard deviation (SD) and coefficient of variation (% CV) were calculated for each sample, for each primer. The geometric mean of the Ct's for RPS18 and β-Actin was calculated and used as the invariant control for ΔCt calculations.

To calculate the % CV and consequently the standard deviation of the ΔCt and the ΔΔCt, the standard deviation of the component Cts was propagated by summation in quadrature[498]. Therefore the %CV for ΔCt for the two genes 1 and 2 becomes;

$$\%CV\ of\ \Delta Ct = \sqrt{\left(\frac{\sigma Ct_1}{Mean(Ct_1)}\right)^2 + \left(\frac{\sigma Ct_2}{Mean(Ct_2)}\right)^2}$$

Equation 2. Summation in Quadrature. σ = standard deviation

2.5.4 RNA interference in Cell Lines

Cells were seeded in standard tissue culture plates in complete medium and allowed to adhere overnight before exchanging the medium for antibiotic-free, supplemented (CP-As) or serum-containing medium (all other cells).

For the optimisation of transfection, Opti-MEM reduced serum medium + Glutamax (Invitrogen) alone or mixed with Dharmafect 1 (ThermoScientific), Dharmafect 4 (ThermoScientific), Lipofectamine 2000 (Invitrogen) or RNAiMAX (Invitrogen) were added to cells growing in antibiotic-free, serum or supplement-containing medium in 6-well plates at 50% confluence. Viability was monitored visually by daily light microscopy for 72 hours post-transfection.

For subsequent experiments using the RNAiMAX protocol, pools of 4 pre-designed siRNAs to genes of interest (Table 2-6) or a non-targeting, scrambled siRNA sequence (siGENOME, Non-Targeting siRNA#3, Dharmacon) were mixed in Opti-MEM reduced serum medium + Glutamax (Invitrogen) and allowed to form complexes for 20 minutes at RT. Complexes were added to the cell culture medium, gently mixed by pipetting (final siRNA concentration 25 nM, RNAiMAX final concentration 0.2% v/v) and cells incubated for a further 48 hours before harvesting for protein or RNA assays. Dharmacon SMARTPools have previously been used successfully in oesophageal cancer cells[355].

To assess the effect of siRNA on cell viability, cells were seeded in poly-L-Lysine coated 96-well plates (CPA, OE33 - 5,000 cells/well; OE19 - 10,000 cells/well, final volume 100µL) and allowed to adhere for 24 hours. A baseline measurement of viability was performed using the AlamarBlue method (2.4.2) and then the standard siRNA transfection protocol followed, as above. A final AlamarBlue measurement was taken 96 hours after siRNA transfection.

For transfection of Accell siRNA, cells were grown in standard complete medium in 6 cm plates and a pool of 4 Accell siRNAs specific to the target gene (Table 2-6 siRNA sequences) or a non-targeting scrambled sequence (Accell non-targeting control #1, Dharmacon) were diluted in complete medium from 100 µM stock and

added to the plates (final concentration; 1 μM). Cells were harvested 48 hours after transfection for assay of RNA or protein expression. For viability assessment, measurements were made using the AlamarBlue method before transfection and 96 hours after transfection in 96-well plates. Transfection efficiency was assessed using FAM-labelled, non-targeting, scrambled sequence Accell siRNA (excitation 494 nm, emission 520 nm) and blue/UV light, bright-field microscopy with images acquired on a standard digital camera with a microscope attachment (Nikon,UK).

Table 2-5 qRT-PCR Primer and Probe Sequences for Solaris Assays.

Target	5′ to 3′ Forward primer	5′ to 3′ Reverse primer	Probe	Amplicon size base pairs (bp)
PLK-1	CAACGGCAGCGTGCAGAT	GCACAAGATGAGCTTGGT	ACTTCTTCCAGGATCACACC	55 bp
TP53	CTCAGCATCTTATCCGAGTGGAAG	TGGTACAGTCAGAGCCAACCTCA	CCCTATGAGCCGCCTGA	120 bp
AURKA	GCAAATGCCCTGTCTTACTG	AGAGTGGTCCTCCTGGAA	TACATGCTCCATCTTCC	146 bp
CDKN1A	GCGACTGTGATGCGCTAA	CCTCCAGTGGTGTCTCGGT	GAACTTCGACTTTGTCACCG	87 bp
GRP-78	GAAGACTTTTCTGAGACCCTGAC	GCTTCATAGTAGACCGGAAC	GCTCAACATGGATCTGT	76 bp
OLFM-4	ATCGTGGCTCTGAAGACCAA	ATGACCACAGCTCCCTGGAGT	CCACCCTCCTCCCACTC	99 bp
GSTO1	ACCTGTGAGTACCTGGATGA	AGCTTCCTACCAAGGATG	GTTTTCTAAGGTGCCATC	124 bp
FBP1	TCCAACGACCTGGTTATGAAC	AGACCACATATTTACCCCTT	CATAGTGGAACCGGAGAAA	118 bp
SAMHD1	CGCAACTCTTTACACCGTAG	TTTGAGGAAAGCATCTGT	TGATACAATGATTACAG	81 bp
GAPDH	GCCTCAAGATCATCAGCAATG	CTTCCACGATACCAAAGTTGTC	GCCAAGGTCATCCATGA	90 bp
RPS18	GTGATCCCTGAAAAGTTCCAG	CACATGAGCATATCTTCGG	CCATTAAGGGTGTGGGC	120 bp
β-Actin	TGGAGAAAATCTGGCACCAC	ACCGCGAGAAGATGACC	GGTCTCAAACATGATCTGG	79 bp

Table 2-6 siRNA sequences

Target	Source	siRNA Sequences
SAMHD1	siGENOME, SMARTPool, Dharmacon	GGAACUCCAUCCCGACUAC, CGUAAUCUAUUCAAGUAUG, UUAGUUAUAUCCAGCGAUU, GACAAUGAGUUGCGUAUUU
GRP-78	siGENOME, SMARTPool, Dharmacon	CCACCAAGAUGCUGACAUU, GAAAGGAUGGUUAAUGAUG, CAGAUGAAGCUGUAGCGUA, CGACUCGAAUUCCAAAGAU
FBP1	siGENOME, SMARTPool, Dharmacon	AGAUAAACACGCCAUCAUA, UGACAGGUGAUCAAGUUAA, GGGCAUCGCGCACCUCUAU, GGGUAAAUAUGUGGUCUGU
AGR2	siGENOME, SMARTPool, Dharmacon	GAAGAAAGCUCUCAAGUUG, GAAAUUGGCAGAGCAGUUU, CAGUCAAGCUUUAAAGAAA, AGUCAAACCUGGAGCCAAA
GSTO1	siGENOME, SMARTPool, Dharmacon	GUAGGAAGCUUUAUUAGAA, GCAAUGAAGUUAAAUGAGU, GAGAAAGACUGGCAAGGUU, UAGAGGAGGUUCUGACUAA
Rho GDI-2	siGENOME, SMARTPool, Dharmacon	GAAAGUGGAUAAAGCAACA, GGAAGGUUCUGAAUAUAGA, GAUGAGAGUCUAAUUAAGU, GAGUGGAACCUGUCGAUUA
PLK-1	siGENOME, SMARTPool, Dharmacon	CAACCAAAGUCGAAUAUGA, CAAGAAGAAUGAAUACAGU, GAAGAUGUCCAUGGAAAUA, CAACACGCCUCAUCCUCUA
TP53	siGENOME, SMARTPool, Dharmacon	GCUUCGAGAUGUUCCGAGA, GAAGAAACCACUGGAUGGA, GCACAGAGGAAGAGAAUCU, GAGGUUGGCUCUGACUGUA
PLK-1	Accell, SMARTPool, Dharmacon	CCAUAUGAAUUGUACAGAA, CCAAGGUUUUCGAUUGCUC, CUGUCCUUUCCUUGGCUUU, GUGGGACUCCUAAUUACA
GAPDH	Accell, Dharmacon	Proprietary Control Sequence
GAPDH	siGENOME, Dharmacon	Proprietary Control Sequence

2.5.5 Statistical Comparisons of Viability and Gene Expression

The mean viability or gene expression level between different conditions was assessed for statistical significance using a two-tailed, un-paired, Student's t-test (GraphPad Prism version 5.00 for Windows, GraphPad Software, San Diego, California, USA). Significance was defined as $p<0.05$.

2.6 Cell Cycle Analysis

2.6.1 Cell Synchronisation, Treatment and Fixation

Cells were seeded in 10 cm plates and grown in complete medium until ~50% confluent. To synchronise cells in G_0 phase, complete medium was exchanged for serum or supplement-free medium and plates incubated for 48 hours before exchange for complete medium supplemented with either PLK-1A or BI2536, both diluted from a 10 mM stock made up in DMSO or DMSO alone. At 24 hours after drug or DMSO treatment, cells were harvested by trypsinisation, resuspended in 10 ml of PBS and 500,000 cells (OE33, OE19) or 750,000 cells (CP-A, HEEpiC) pelleted by centrifugation for 5 minutes at 400 g and resuspended in 300 µL PBS. Cells were fixed by drop-wise addition of 900 µL of ice-cold 70% (v/v) Ethanol in dH_2O while vortexing. Fixed cells were stored at 4^{o}C for up to 7 days before Propidium Iodide (PI) staining.

2.6.2 Propidium Iodide Staining

Fixed cells were centrifuged for 5 minutes at 400 g and then resuspended in 5 ml of PBS before centrifugation again for 5 minutes at 400 g. The supernatant was discarded and cell pellets suspended in 300 µL Propidium Iodide staining solution (50 µg/ml Propidium Iodide and 100 µg/ml RNAse in PBS) and incubated at RT in the dark for 60 minutes.

2.6.3 Flow Cytometry

Propidium Iodide stained cells were injected into a BD Aria II Flow Cytometer with cellular fluorescence (excitation 488 nm, emission 635nm) recorded from forward and side-scatter detectors and manually gated using FACSDiva software version 6.1.3 (BD Biosciences). A minimum of 10,000 events were recorded for each sample. Baseline cellular DNA content was assumed to present a diploid genome (2N) and the percentage of 4N and 2N cells calculated. Cells with DNA content between 2N and 4N peaks were assumed to represent the S-phase fraction.

2.7 Immunohistochemistry

2.7.1 OCCAMS TMA

Archival formalin-fixed paraffin-embedded (FFPE) tissue blocks from oesophageal adenocarcinoma resection specimens were obtained from oesophago-gastric resection centres in Bristol, Birmingham, Papworth, Oxford and Glasgow as part of the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS) collaborative network and used to create a tissue microarray (TMA) as described previously[275]. Briefly, cores were cut using 1 mm punches from areas confirmed on Haematoxylin and Eosin stained sections to contain adenocarcinoma, normal squamous epithelium or Barrett's oesophagus. Cores were then embedded in paraffin blocks to construct the array. For evaluation of p53 and PLK-1 staining, slides from the OCCAMS tissue microarray were made available for this project via a collaborative agreement with the OCCAMS Chief Investigator, Prof. Rebecca Fitzgerald (Cambridge).

2.7.2 Immunohistochemistry (IHC) for PLK-1 and p53

Sections (4 µm) were cut from the OCCAMS TMA blocks using a microtome, mounted on slides (VWR superfrost) and dehydrated at 56^{o}C overnight before staining. Antibodies were only used for IHC if target specificity could be confirmed. Staining for p53 and PLK-1 was performed by Mr. Johnny Ong (Cambridge) using a BOND automated immunostainer (Leica Microsystems). This system combines the dewaxing, rehydration, antigen retrieval, blocking, primary and secondary antibody incubation steps in an automated workflow to allow efficient and reproducible sample processing. For Bond automated staining, the following settings were used;

Target	Antibody Source, Product Number, Dilution	Automated IHC Programme
p53	Leica, NCL- L-p53-DO7, 1:50 dilution	IHC Protocol F + 3,3'-Diaminobenzidine (DAB) Enhancer. Heat Induced Epitope Retrieval Solution 1 (Citrate buffer at pH 6.0) for 30 minutes.
PLK-1	Upstate, #05-844 (clone 35-206), 1:7500 dilution	IHC Protocol F + DAB Enhancer. Heat Induced Epitope Retrieval Solution 2 (EDTA based buffer pH 9.0) for 30 minutes

2.7.3 IHC for Other Proteins

For other targets, IHC was performed in Edinburgh using the following method.

2.7.4 Optimisation of IHC conditions

Formalin fixed tissue, surplus to diagnostic requirements was obtained from the resection specimens of prospectively-consented patients undergoing resection for oesophageal cancer (Ethical Approvals; 06/S1101/16, 10/S1402/33, NHS R&D Approvals; 2006/W/PA/01, see 2.8). Tissue representing normal oesophagus, normal stomach and oesophageal adenocarcinoma were sampled from the oesophagogastric resection specimen by a consultant histopathologist. Tissue pieces were then embedded in paraffin and used for optimization of IHC staining.

Sections (4 µM) were cut by microtome from FFPE blocks and mounted on electrostatically-coated slides (Superfrost, VWR). Slides were dehydrated at 56^{o}C overnight before dewaxing and rehydration by incubation at RT in a series of solvents (Table 2-7).

Table 2-7 Dewaxing and Rehydration of FFPE sections.

Solvent (In Order of Incubation)	**Incubation Time (Room Temperature)**
Xylene	2 x 5 minutes
Ethanol (Absolute)	2 x 2 minutes
Ethanol (90% v/v in dH_2O)	2 minutes
Ethanol (70% v/v in dH_2O)	2 minutes
H_2O	Until Antigen Retrieval (maximum 60 minutes)

Antigen retrieval and primary antibody dilutions were optimized using sections from OAC tissue and a "checkerboard" titration of conditions (Table 2-8). Antigen retrieval was performed in citric acid buffer (1.8 mM Citric Acid, 8.2 mM Sodium Citrate in dH_20, pH 6.0) in 50 mL Coplin jars incubated in a water bath set to 101^{0}C or in a pressure cooker filled with 1000 ml citric acid buffer, pre-heated to 95^{o}C and then microwaved at 1000W power for 5 minutes. Slides were left to cool in antigen retrieval buffer for 20 minutes.

Table 2-8 Example Checkerboard Titration

Primary Antibody Dilution	No Antigen Retrieval	Citrate Buffer (pH 6.0) for 20 minutes water bath @95°C	Citrate Buffer (pH 6.0) for 40 minutes water bath @95°C	Citrate Buffer (pH 6.0) for 5 minutes pressure cooker.
1/50				
1/100				
1/200				
No Primary				

Endogenous tissue peroxidase activity was blocked by incubation in 3% (v/v) Hydrogen peroxide (H_2O_2) in dH_2O for 10 minutes with stirring, followed by rinsing in water. If initial conditions produced prominent background staining, an additional blocking step was added using 200 µL of Serum-Free Protein Block (SFPB, DAKO) per slide for 20 minutes at RT.

Blocked slides were incubated with primary antibody in antibody diluent (DAKO) for 1 hr at RT (200 µL per slide) followed by TBS for 5 minutes and incubation with undiluted Envision secondary antibody solution (DAKO) for 30 minutes at RT. Slides were again washed in TBS and then incubated in 3,3'-Diaminobenzidine chromogen (DAB, DAKO) for 10 minutes at RT before rinsing in water.

Slides were all counter-stained in Haematoxylin solution (5% w/v Aluminium Potassium Sulphate, 1% Haematoxylin w/v, 0.2% w/v Sodium iodate, 2% v/v Glacial Acetic Acid) for 30 seconds, rinsed in tap water , blued in Scott's Tap Water substitute (Leica) for 30 seconds and rinsed again in tap water. Slides were then finally dehydrated in a series of graded solvents (Table 2-9).

Table 2-9 Dehydration of stained FFPE sections.

Solvent (In Order of Incubation)	**Incubation Time (Room Temperature)**
Ethanol (70% v/v in dH_2O)	1 minute
Ethanol (90% v/v in dH_2O)	1 minute
Ethanol (Absolute)	2 x 1 minute
Xylene	2 x 1 minute

Coverslips (Fisher) were mounted onto slides using DPX mountant and the staining pattern was scored after drying using the following criteria;

Intensity of positive (brown) staining - 3 points for maximum intensity to 0 points for no positive staining.

Lack of background staining - 3 points for no background staining to 0 points for intense widespread background staining.

Relative positive staining to background ratio – Slides with the strongest positive staining may also feature non-specific background staining that clouds interpretation. A compromise between strength of positive staining and strength of background staining is often needed to provide the clearest picture. This was scored from 3 points for good demonstration with no non-specific staining, to 0 points for no target-specific positivity but strong widespread non-specific staining.

Scores were tabulated across the checkerboard titration. Conditions with the highest scores were further optimised by the addition of blocking stages, change of antigen retrieval conditions or primary antibody dilutions. Once optimal conditions were achieved for each antibody using OAC tissue, the staining pattern across normal oesophagus and normal stomach was examined and conditions further optimized until a consistent pattern was observed on at least 2 independent occasions. The optimised conditions used for candidate staining are summarised in (Table 2-10). For acquisition of images from stained slides, a digital slide scanner was used (Ariol, Leica).

2.7.5 Confirmation of Expression in FFPE Tissue

For the confirmation of protein expression by IHC, two sections (4 µm) were cut from blocks representing normal oesophagus, normal stomach, adenocarcinoma, lymph nodes involved with metastatic adenocarcinoma (if present) and uninvolved nodes. One slide was stained using the appropriate primary antibody and an optimized protocol (Table 2-10) and a no primary control slide was included for each protein.

Table 2-10 Optimised IHC conditions

Antibody, Source, Reference	Antigen Retrieval (citric acid pH 6.0)	Blocking Conditions	Primary dilution of manufacturer stock	Positive control tissue
AGR2, Sigma, HPA007912	20 mins WB	H_2O_2 10 mins then SFPB 20 mins	1/1000 in diluent	ACC + Barrett's
SAMHD1, Abcam, ab119751	5 mins PC	H_2O_2 10 mins then SFPB 20 mins	1/600 in diluent	ACC
GRP78, Abcam ab21685	20 mins WB	H_2O_2 10 mins then SFPB 20 mins	1/200 in diluent	ACC
Rho-GDI-2, Abcam ab88317	20 mins WB	H_2O_2 10 mins then SFPB 20 mins	1/600 in diluent	Inflammatory cells in Stroma
HSP27, Abcam ab114067	No AR,	H_2O_2 10 mins then SFPB 20 mins	1/1000 in diluent	Squamous epithelium
TGM3, Sigma, HPA004728	5 mins PC	H_2O_2 10 mins then SFPB 20 mins	1/300 in diluent	Squamous epithelium
WB – Water Bath, PC – Pressure Cooker, AR – Antigen Retrieval, SFPB – Serum Free Protein Block,				

The staining pattern of each tissue component on the slides was tabulated and slides reviewed independently by a Consultant Upper GI Pathologist (Dr. Vicki Save, Edinburgh) to confirm the accuracy of scoring.

2.7.6 Scoring of TMA Staining

Stained slides from the TMA were scanned using a digital slide archival system (MIRAX SCAN, Zeiss, Cambridge UK) and images scored using the Allred scoring system[499]. Using this method, two scores are generated for each core, one for the relative intensity of staining and one for the approximate proportion of tumour cells staining (Table 2-11). For p53 IHC scoring, cores were also scored with a predicted genotype of wt, null or mutant.

Table 2-11 Allred Scoring System Adapted for Oesophageal TMA Scoring

Proportion Score		+	Intensity Score		=	Total Score (Range 0,2-8)
Proportion of Tumour Cells Staining	Score		Staining Intensity	Score		
0%	0		Nil	0		
~1%	1		Low	1		
~10%	2		Medium	2		
~33%	3		High	3		
~66%	4					
~100%	5					

Total Allred histoscores for PLK-1 and p53 staining and p53 genotypes from two independent scorers (this author and Dr. Rudolf Nenutil, Brno, Czech Republic) were tabulated for each slide and TMA maps subsequently used to associate cores from the same tumour together.

2.7.7 Confirmation of IHC Scoring Pattern and p53 Genotype

The accuracy of p53 allele prediction from the IHC staining pattern was assessed on a small sample of tumours undergoing whole-genome sequencing as part of the International Cancer Genome Consortium (ICGC) Oesophageal Cancer Genome Project[500]. The p53 allele status was determined for each of these tumours as part of the standard mutation calling algorithm and verified by targeted amplicon resequencing (Prof. Rebecca Fitzgerald, manuscript submitted).

2.7.8 Assessment of Inter-observer Agreement

Both scorers independently assessed the same images and provided predicted p53 genotypes and Allred histoscores for all cores on the TMA slides. The Kappa coefficient was used to assess the degree of interobserver agreement for p53 genotypes[501]. A kappa of 1.0 indicates perfect inter-observer agreement and 0 indicates no agreement above that expected by chance. An example calculation is provided for the theoretical example in Table 2-12.

Table 2-12 Two Observer Example for p53 Genotype Prediction

		Observer 2		
		Wildtype	Mutant	Null
Observer 1	Wildtype	a	b	c
	Mutant	d	e	f
	Null	g	h	i

$$p(Agreement) = \frac{(a + e + i)}{t}$$

$$t = (a + b + c + d + e + f + g + h + i)$$

$$p(Chance) = \left(\frac{(a + d + g)}{t} * \frac{(a + b + c)}{t}\right) + \left(\frac{(b + e + h)}{t} * \frac{(d + e + f)}{t}\right) + \left(\frac{(c + f + i)}{t} * \frac{(g + h + i)}{t}\right)$$

The kappa value is calculated using the following formula;

$$Kappa = \frac{p(Agreement) - p(Chance)}{1 - p(Chance)}$$

Equation 3 Calculation of Kappa

For assessment of the agreement of inter-observer histoscores for each core and median histoscores from the three cores for each tumour, the Spearman correlation coefficient was calculated using GraphPad Prism (version 5.00 for Windows, GraphPad Software, San Diego, California, USA).

2.7.9 Ovarian Cancer TMA staining

A tissue microarray created from 99 ovarian cancers comprising serous, mucinous and endometrioid adenocarcinomas with known p53 genotypes was subjected to PLK-1 IHC by Dr. Rudolf Nenutil (Brno, Czech Republic). Five patients had undergone neoadjuvant chemotherapy and the remainder were chemotherapy naive. The same staining protocol was used as for the oesophageal TMA. Scoring was

undertaken by a single observer (Dr. Nenutil), blinded to the p53 genotype of the tumour.

2.7.10 Comparison of Median Histoscores

Median tumour histoscores were compared for groups defined by predicted or actual TP53 genotypes using the Kruskal-Wallis test of significance calculated using GraphPad Prism (version 5.00 for Windows, GraphPad Software, San Diego, California, USA).

2.8 Prospective Collection of Tissue, Blood and Data

2.8.1 Regulatory Approvals

An Oesophageal Tissue Bank (EOTB) was created to facilitate translational research. This was established under the wider framework of the Edinburgh Experimental Cancer Medicine Centre (ECMC). Prospectively obtained ethical (REC references 06/S1101/16 and 10/S1402/33) and NHS Lothian Research and Development (R&D ID 2006/W/PA/01) approvals were in place (see Appendix). Under these approvals, fresh frozen tissue, surplus to diagnostic requirements, was collected from consenting patients undergoing attempted curative therapy for OAC.

2.8.2 Patient Recruitment

The identification and recruitment of eligible patients was integrated into the weekly South-East Scotland Oesophagogastric Cancer Multidisciplinary meeting. Standard operating procedures (SOPs) outlining the process of patient eligibility determination, consent, blood and tissue collection are provided in the Appendix.

In brief, patients were approached after the diagnosis of localized oesophageal adenocarcinoma and the decision to pursue attempted curative therapy incorporating surgical resection. After outline of the process of tissue and blood collection, informed consent was obtained. Blood was collected and processed according to the SOP and de-identified clinical data obtained from patient records and stored on a bespoke secure database. The integration of clinical data, tissue and blood collection into the standard clinical care pathway is summarised in Figure 2-1.

2.8.3 Tissue collection

At the time of surgery, resection specimens were obtained fresh and transported to the pathology department. Biopsies representing macroscopically normal oesophagus and stomach along with adenocarcinoma tissue were collected from resection specimens by a Consultant Upper-GI pathologist. Biopsies were collected directly into low binding cryovials and snap frozen and maintained in liquid nitrogen in a storage facility or on dry ice during transport and as long as possible up until use to minimize proteolytic degradation.

2.9 Proteomics

2.9.1 Tissue Handling

All preparation of tissue samples for proteomic studies was performed using keratin-free consumables including low-residue keratin-free tips (Maxymum Recovery tips, Thistle Scientific) with standard precautions to minimize keratin contamination. Biopsies were maintained in liquid nitrogen or on dry ice until lysis. All buffers were prepared in positive pressure fume-hoods.

2.9.2 Protein Extraction from Human Tissue Biopsies-Pilot Study

For the pilot study, a protocol previously applied to breast cancer tissue was adapted for use with oesophageal biopsies[431]. Pieces of snap frozen oesophageal tissue, 4-5 mm^3 in size were placed in 2 ml, autoclaved, low-binding micro-centrifuge tubes (Axygen, Maxymum Recovery, Corning) containing sterile 5 mm steel beads. Lysis buffer (80-150 µL, 500 mM triethylammonium bicarbonate (TEAB), pH 8.5 and 0.05% (v/v) SDS, both mass spectrometry (MS) grade) was added to the tubes in a volume proportional to biopsy size. Biopsies were then immediately homogenized using a bead-mill homogeniser (Retsch, Qiagen) for 2 cycles each of 2 minutes at 25 Hz. The homogenate was transferred to a fresh low-binding 1.5 ml tube (Maxymum recovery, Thistle Scientific) and kept on ice before sonication using a needle sonicator at a power of 25W, with the cycle set at 30 x 0.1 s, 30 s pause and 30 x 0.1 s. Sonicated samples were left for 75 minutes on ice and then centrifuged at 14,000 g, for 20 mins at 4^{o}C. Supernatants (lysate) were snap-frozen and stored in liquid nitrogen until use.

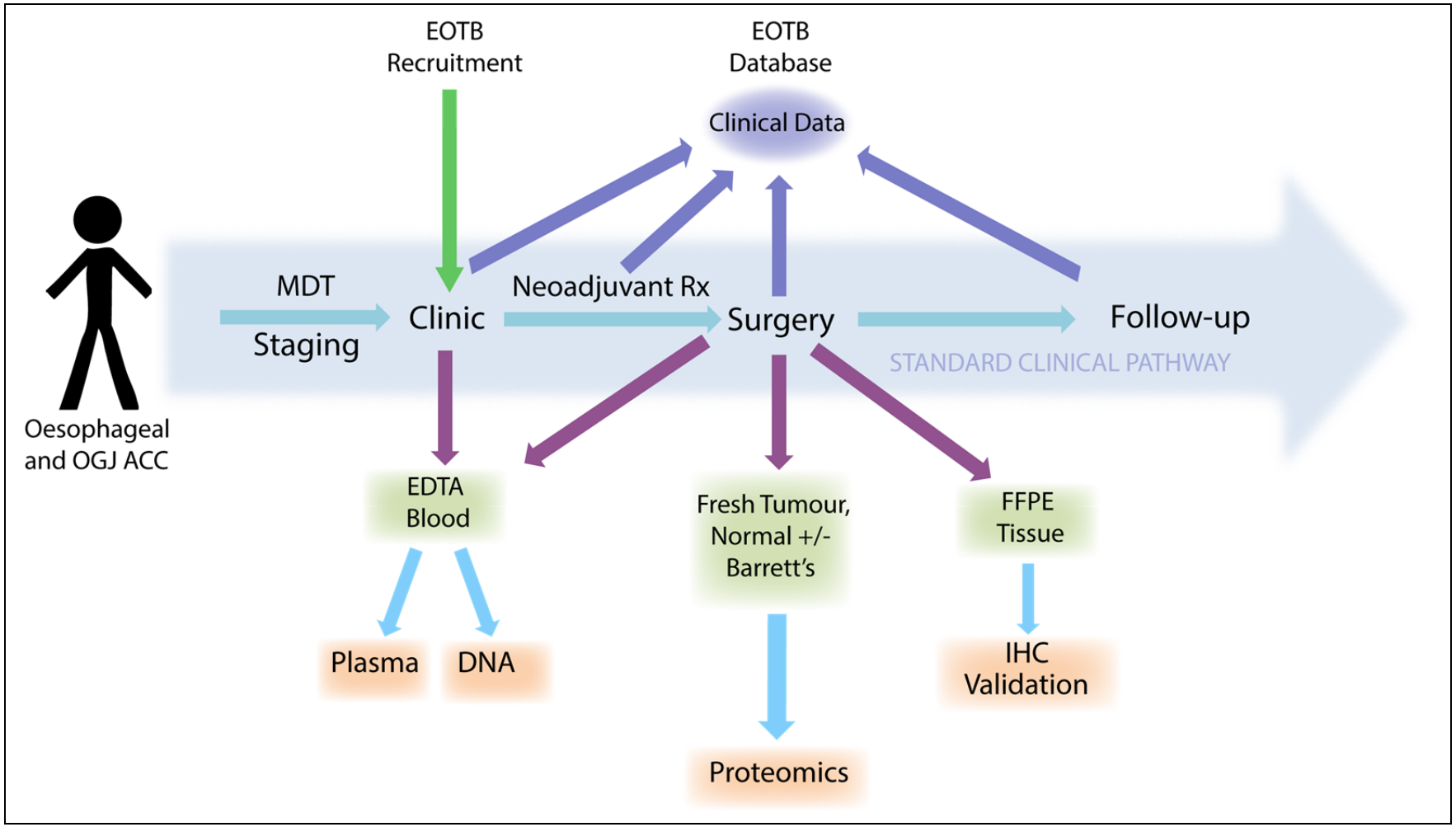

Figure 2-1 Edinburgh Oesophageal Tissue Bank (EOTB) Pathway. Patients staged with localised OAC, undergoing attempted curative treatment were approached for consent for tissue and blood donation of research purposes. The integration of tissue and data collection into the clinical care pathway is summarised. MDT – Multidisciplinary Team, Rx - Treatment

2.9.3 Protein Extraction from Human Tissue Biopsies – Subsequent Studies

To enhance proteome coverage, a modified method of tissue lysis was developed based on the published Filter-Aided Sample Preparation (FASP) protocol[443]. This method has two key features; tissues are lysed in a buffer containing a high detergent concentration to improve protein solubilisation, and then the buffer is exchanged to a trypsin and MS-compatible solution by centrifugation of lysates through molecular weight cut-off filters.

2.9.4 Optimisation of Tissue Lysis Conditions.

2.9.4.1 Tissue Handling

Samples of mouse small intestine were collected from 12 mice culled after pentobarbital anaesthesia as part of an independent, ethically-approved research project (Courtesy of Dr. Chris Lucas, Queen's Medical Research Institute, Edinburgh). Biopsies were snap-frozen on dry ice and weighed by subtraction into 2 ml homogenisation tubes (Matrix A or Matrix D, MP Bio).

2.9.4.2 Homogenisation

Detergent-free lysis buffer (DFLB, 100 mM Tris- HCL, 100 mM DTT, pH 7.6), Complete buffer (SDSB, 100 mM Tris- HCL, 100 mM DTT, 4% (v/v) SDS, pH 7.6) or no buffer were added to the tubes prior to homogenisation at RT by rapid shaking in a bench-top homogenizer (FastPrep-24, MP Bio) for 40 seconds at 6 ms^{-1}. Tubes were then centrifuged for 15 s at 13,000 rpm in a bench-top centrifuge to re-collect homogenised material. Complete buffer, SDS or no buffer were added to the tubes to complete the lysis. Tubes were then gently mixed by rotation for 20 minutes at RT before sonication or proceeding directly to buffer exchange.

2.9.4.3 Sonication

Lysates were sonicated on ice using a needle sonicator (Bioruptor) for 30 seconds at maximum amplitude. Sonicated lysates were heated for 5 minutes at 95^{o}C and then transferred to low-binding micro-centrifuge tubes (Eppendorf) before centrifugation at 14,000 g for 5 minutes at 20^{o}C.

2.9.4.4 Buffer Exchange

Clarified lysates were diluted in 200 µL urea solution (8 M urea in 100 mM Tris HCl, pH 8.5) in 30 kDa nominal molecular weight cut-off filter (NMWCO, Millipore; 30µL lysate per filter) and then centrifuged for 15 minutes at 14,000 g at 20°C. The residuates were re-diluted by the addition of a further 200 µL of urea solution and again centrifuged for 15 minutes at 14,000 g at 20°C. For optimization experiments, a 25 µL aliquot of buffer exchanged residuate was used for protein concentration determination (see 2.9.6).

2.9.4.5 Lysate Alkylation and Urea Dilution

For each filter, the residuate (buffer-exchanged lysate) was mixed with 100 µL of 50 mM iodoacetamide in urea solution and incubated in the dark at RT for 20 minutes. Alkylated lysates were then centrifuged at 14,000 g for 10 minutes at 20°C and again diluted by the addition of 100 µL of urea solution before centrifugation again at the same settings for 15 minutes. This step was repeated twice and then residuates mixed with 100 µL of 50 mM triethylammonium bicarbonate (TEAB) to dilute the urea and centrifuged at 14,000 g for 10 minutes at 20°C, repeating this step twice.

2.9.4.6 Tryptic Digestion

The residuate from one identically treated filter column was used to determine the protein concentration and the remaining filter columns used for on column tryptic digestion. Vials of lyophilised trypsin (20 µg/vial, Promega) were reconstituted in 10 µL of 1 M HCl and then diluted to 20 µg/ml in 20 mM TEAB and 9% (v/v) acetonitrile in dH_2O. Trypsin was added to buffer-exchanged lysates in a ratio of 100:1 (lysate protein mass (mg) : trypsin mass (mg)). Filter units were sealed with Parafilm (Camlab) and incubated overnight at 37°C.

2.9.5 Application of the Modified FASP protocol to Human Tissue

Pieces of snap-frozen oesophageal or gastric tissue between 30mg and 60mg in weight were disrupted at RT in low-binding micro-centrifuge tubes containing 1mm ceramic beads (Matrix D, MP Bio) by rapid shaking in a bench-top homogenizer (FastPrep-24, MP Bio) for 40 seconds at $6ms^{-1}$. The homogenisation step was repeated once if the tissue was not completely disrupted. Homogenates were

dissolved in FASP lysis buffer (4% w/v SDS, 100 mM Tris/HCl, 100 mM DTT, pH 7.6) at a ratio of 20 : 1; (lysis buffer (µL) : tissue weight (mg)) by re-pipetting and then gently mixed by rotation for 20 minutes at RT.

Sonication, lysate clarification, buffer exchange, alkylation and dilution were performed as previously (see 2.9.4.3 and 2.9.4.5) with several lysate aliquots buffer-exchanged for each tissue sample. The NMWCO filter residuates from the same tissue samples were aspirated from the filter units and pooled for protein concentration determination.

2.9.6 Determination of Protein Concentration from Tissue Lysates.

Lysate protein concentrations were determined using a modified Lowry procedure[502] (RC-DC, Bio-Rad) as per manufacturer's recommendations. A dilution series from 4 mg/ml to 0.125 mg/ml of BSA in 50 mM TEAB was used to generate a standard line by linear regression of background-corrected absorbances. Sample protein concentrations were then inferred by interpolation of background corrected absorbances on the standard regression line.

2.9.7 Preparation of Tryptic Peptides from Lysates – Pilot Study.

Each tissue lysate was diluted to 5 µg/ µL in lysis buffer (500 mM TEAB (pH 8.5) and 0.05% (v/v) SDS) and then an 8µ L aliquot diluted with 32 µL of 6 M urea in 100 mM TEAB (pH 8.0). Proteins were then reduced for 1 h at 37°C by the addition of 2 µl 50 mM tris-(2-carboxyethyl) phosphine hydrochloride (TCEP) in dH_2O. The reduced lysates were alkylated by the addition of 1 µL Iodoacetamide (400 mM) and incubated for 30 minutes at RT, protected from light. Lysates were then further diluted with 67 µL of 100 mM TEAB and digested with trypsin (MS grade, Promega; final concentration 0.167 µg/µL, protein : enzyme ratio of 20 : 1) overnight at 37^{o}C.

2.9.8 Preparation of Tryptic Peptides from Lysates – Modified FASP method

Vials of lyophilised trypsin were reconstituted as previously and added to pooled lysates in a ratio of 100:1 (Lysate protein mass (mg) : Trypsin mass (mg)). The pH of the final mixture was confirmed to be between 7.5 and 8.5. Eppendorf tubes were sealed with Parafilm and tryptic digestion conducted overnight at 37^{o}C.

2.10 Quantitative Proteomics using Isobaric Labeling

2.10.1 Isobaric Labeling of Tryptic Peptides

Tryptic peptide mixtures from each tissue sample were independently labeled with one of the 6 TMT reagents (ThermoScientific) according to manufacturer's instructions. For the pilot study, two equal aliquots containing a mixture of all 4 samples were labeled with the two remaining tags as an internal control to define the level of technical variation. For further experiments, each tissue type from one patient was labeled in technical duplicate. Labeled samples were pooled and then evaporated to dryness in a vacuum centrifuge (SpeedVac, ThermoScientific).

Peptides were labelled with TMT reporters, fractionated by OFFGEL electrophoresis, and analysed by liquid chromatography - tandem mass spectrometry (LC - MS/MS) by Dr. Alex Scherl and Mr. HuiSong Pak (Centre Médical Universitaire, Geneva) on a linear ion trap-Orbitrap hybrid mass spectrometer (LTQ-Orbitrap Velos, ThermoScientific) equipped with an ultra-high pressure, liquid chromatography (UPLC) system (nanoACQUITY, Waters).

2.10.2 Peptide Fractionation for TMT Experiments

TMT-labeled peptides were reconstituted in 200 µL dH_2O (MS grade, Sigma) and desalted on a Macro SpinColumn C18 (Harvard Apparatus) before evaporation by vacuum centrifugation. Peptides were reconstituted in 1 x OFFGEL solution (a proprietary solution containing a mixture of urea, thiourea, DTT, glycerol and Tris-based buffer, Agilent) and applied to an immobilised pH gradient (IPG) dry strip (13 cm, pH 3–10, linear; GE Healthcare) set up with a 12-well frame for the pilot study and 24-well frame for subsequent experiments. Peptides were then fractionated according to their isoelectric point by OFFGEL electrophoresis with a maximum

current of 50 µA and power of 200 mW until 20 Vh were reached (3100 OFFGEL Fractionator, Agilent)[503]. The fractions were collected and desalted on a Micro SpinColumn C18 (Harvard Apparatus) before evaporation to dryness under speed-vacuum and storage at -20 °C until LC-MS/MS.

2.10.3 Liquid Chromatography for TMT Experiments

Dried, desalted peptides were reconstituted in 5% (v/v) acetonitrile, 0.1% (v/v) Formic Acid (FA) in dH_2O and ~0.5 µg loaded onto a homemade 100 µm internal diameter, 20 mm long trapping column packed with 200 Å, 5 µm Magic C18 AQ (Michrom, Auburn, CA, USA). Trapped peptides were eluted into a 750 µm internal diameter, 150 mm long analytical column packed with 100 Å, 5 µm Magic C18 AQ (Microcom).

For UPLC, peptides were separated using a variable solvent gradient created by a combination of 0.1% (v/v) FA in H_2O (solvent A) and 0.1% (v/v) FA in acetonitrile (solvent B). The gradient was run as follows: 0–1 mins, 95% (A) and 5% (B), 1-56 mins, 65% (A) and 35% (B), 66-76 mins, 20% (A) and 80% (B) using a flow rate of 220 nL/min.

2.10.4 Mass Spectrometry for TMT Experiments

Peptides were analysed in positive ion mode after electrospray ionisation. For MS survey scans, the OT resolution was set to 60,000 and the ion population was set to 5×10^5 with an m/z window from 400 to 2,000.

For precursor ion analysis (MS1), a maximum of 3 ions with the greatest peak intensities were selected for both collision-induced dissociation (CID) and high-energy C-trap dissociation (HCD) in the LTQ with analysis in the OT. For fragment ion analysis (MS2) in the LTQ, the ion population was set to 7×10^3 (isolation width of 2 m/z) while for MS2 detection in the OT, it was set to 2×10^5 (isolation width of 2.5 m/z), with resolution of 7,500, first mass at m/z = 100, and maximum injection time of 750 ms. The normalized collision energies were set to 35% for CID and 60% for HCD.

2.11 Data analysis for TMT Experiments

2.11.1 Protein Identification

Protein identifications were made using the Easyprot platform (v2.3 build 720, Swiss Institute of Bioinformatics)[504]. Data manipulation was performed using Excel (Version 14.0.6129.5000, Microsoft Office Professional 2010), R (version 2.15.1, General Public License), and custom scripts written in Perl (version 5.18.0, General Public License). Example code is provided in the Appendix.

The Easyprot platform proceeds as follows: peak lists were generated from Thermo RAW files using ReAdW (ThermoFinnigan) and CID and HCD spectra were merged for simultaneous identification and quantification[505]. For the pilot study, the three technical replicates were pooled and peaklist files were searched against the uniprot_sprot database (2011_2 of 8-feb-2011) specifying Homo sapiens taxonomy and using Phenyx peptide spectrum matching software (Swiss Institute of Bioinformatics). For subsequent TMT experiments, peaklist files were searched against the Uniprot Human Reference Proteome (release 09/01/2013, containing 87,613 entries) using Phenyx. This reference database comprises both manually reviewed (UniprotKB/Swissprot) and unreviewed sequences (UniprotKB/TrEMBL).

For database searching, the precursor ion tolerance was set to 10 parts per million (ppm). TMT-modified peptide N-termini and lysines (additional 229.1629 Da) and oxidized methionines were set as variable modifications. Carbamidomethylation of cysteines was set as a fixed modification. Trypsin was selected as the digestion enzyme, with one potential missed cleavage accepted. A minimum peptide length of 6 amino acids was required with a minimum peptide z-score of 4.

All datasets were searched separately, once in a forward and once in a reversed database. A peptide z-score threshold was then calculated to maintain the false positive peptide ratio below 1%. Using Phenyx search scores this approach has been reported to slightly overestimate the true false-positive ratio[506]. For the pilot study, only proteins matching two unique peptide sequences were kept. To maximise proteome coverage for subsequent experiments, a single unique peptide was accepted for protein identification.

2.11.2 TMT Quantitation

2.11.2.1 Reporter Ion Extraction

For relative protein quantitation, the TMT reporter intensities were extracted for each peptide. An isotopic purity correction was performed within Easyprot for each reporter on the basis of the isotopic distribution of the sixplex-TMT reporters provided by the manufacturer. Isotopic purity equations were calculated using Maple software (Maple 11, Maple Inc., Waterloo, Canada).

2.11.2.2 Relative Protein Expression

For the pilot TMT experiment, a tumour to normal oesophageal (TvO) ratio was calculated for each peptide for the first patient as the ratio of 126 to 127 reporter ion intensities and similarly for the second patient as the ratio of 129 to 130 reporter ion intensities. An estimate of technical variation was obtained for each peptide through the ratio of 128 to 131 intensities.

For the subsequent TMT experiments, two strategies were employed. For the first comparison of mFASP TMT data from patients 44 and 46 with PAcIFIC data, the geometric mean of all tumour peptide reporter ions (126 and 127) was calculated for each protein. The same was performed for all normal oesophageal and normal gastric reporters (oesophageal; 128 and 129, gastric; 130 and 131). The tumour : normal oesophageal (TvO) and tumour : normal gastric (TvG) ratios of geometric mean reporters was then calculated for each protein. The peptide level variance was not considered in this initial analysis, as this was unavailable for the comparator spectral counts.

For subsequent comparisons, TvO and TvG ratios were calculated for each peptide as the ratio of 126 to 128 reporter intensities and 126 to 130 reporter intensities respectively. This was repeated for 127/129 (TvO) and 127/131 (TvG) reporter ions as a technical replicate. The geometric mean peptide TvO and TvG ratios were then calculated for each protein to derive an estimate of the relative protein expression between patient-matched tissue types. The geometric mean of ratios was used in preference to the arithmetic mean to limit the skew introduced by outlier ratios[507].

The variance of each ratio and the number of peptides contributing to the mean ratio were included in significance calculations.

Isoforms of some proteins were represented in the reference databases used for peptide spectrum matching. For several genes, more than one protein isoform was potentially detected. The evidence for isoform specific expression could not be empirically determined from the results of peptide spectrum matching and isoform specific protein identifiers could not easily be mapped to unique transcripts. For this reason, the peptides corresponding to different isoforms of a parent protein were pooled and the reporter intensity ratios considered together under the parent protein identifier. The number of unique proteins identified may therefore be underestimated.

2.11.3 Assessment of Technical Variability

The $\log_2$ ratios of technical replicate reporters were calculated for each peptide. Differences from 0 in the median $\log_2$ ratio arise principally due to variations in the initial peptide concentration of samples labelled with each reporter ion. A further small potential contribution could be from differential loss or detection of reporter-labelled peptide during fractionation and LC-MS/MS analysis. This systematic error was minimised by normalisation following the assumption that for each technical replicate ratio the median ($\log_2$ ratio) should be 0.

The variance of the median normalised ratios was used to compare the variability across experiments. By calculating the standard deviation from the combination of all technical replicate ratios, an estimate of the range of ratios expected through technical variation could also be determined.

2.11.3.1 Deriving a Mean Expression Value – Subsequent Experiments

A meta-analysis approach using a fixed-effect model was employed to derive a mean expression value across experiments[508]. The inverse of the variance in peptide reporter ratios was used to weight the contribution of protein expression from a particular experiment to the mean. In this manner, experiments detecting a large

number of peptides all with similar relative tissue expressions for a single protein contribute a greater proportion to the mean expression value.

The weight of a protein expression value (w) for one experiment was defined as;

$$w = \frac{1}{\sigma^2}$$

Equation 4 Weighting used for the contribution of each experiment to the mean. σ^2 = the variance in peptide ratios.

The weighted mean was then calculated as;

$$\overline{X} = \frac{\sum_{i=1}^{k} w_i x_i}{\sum_{i=1}^{k} w_i}$$

Equation 5 Calculation of the Weighted Mean. K=total number of experiments the protein was quantified in. X_i = the expression in experiment i. W_i = the weight for experiment I (see Equation 4).

The variance of the weighted mean was estimated as;

$$\sigma^2 = \frac{1}{\sum_{i=1}^{k} w_i}$$

Equation 6 Variance of the Weighted Mean

The intra-experimental variance associated with a protein expression ratio could not be determined for proteins identified by a single peptide. These proteins may show tissue-specific expression, however, and they were included in the analysis. The un-weighted arithmetic mean and variance of the ratios across experiments was calculated for these proteins.

2.11.3.2 Assessing the Distribution of Ratios

The distributions of the mean, median-normalised, $\log_2$ TvO, TvG, and technical replicates ratios (TvT, OvO, GvG) were examined using histograms and boxplots. To formally test for normality, both the Shapiro-Wilks and Lillefors tests were used[509,510]. To test ratios for homoscedastiscity (equal variances), the Fligner-Killeen test was applied[511].

To determine the false positive performance of the normality tests, random normally distributed variables with the same mean, SD and sample size for the TvO, TvG, TvT, OvO, GvG ratios were generated using the R function Rnorm. A sample of 5000 numbers from these variables selected randomly, with replacement, was tested for normality using both the Shapiro-Wilks and Lillefors tests.

2.11.3.3 Prediction of Dysregulated Proteins at the Population Level

The geometric mean ratio of TvO and TvG peptide reporter intensities was calculated for each protein, and the arithmetic mean calculated across experiments to derive a global mean ratio. The variance of this ratio was determined from biological replicates and the number of biological replicates contributing to the mean used as "n", giving a maximum of 7. To assess technical variation, the TvT and OvO ratios or TvT and GvG ratios were pooled and the arithmetic mean and variance across experiments calculated. The number of biological replicates contributing to each technical replicate ratio was as "n" giving a maximum of 14 for each pooled mean.

2.11.3.4 Calculation of Significantly Dysregulated Proteins

Welch's t-test was used to test the hypothesis that the relative protein expression between tissues was not different from relative protein expression between technical replicates, within the experimental data. This p-value was considered as the probability the protein was not differentially expressed within the experimental data (the intra-experimental probability). The same test was also applied to determine the probability that the relative protein expression between tissues across experiments was not different from the relative protein expression between technical replicates across experiments[512]. This p-value was considered the inter-experimental probability that the protein was not differentially expressed.

For the calculation of intra-experimental probability, the "n" was defined as the number of peptides contributing to the mean ratio. For inter-experimental probability, "n" was defined as the number of patient's samples contributing to the mean ratio. The Welch's t statistic was calculated for each protein for the difference between the mean TvO or TvG ratio and pooled mean TvT and OvO or TvT and GvG ratios (Equation 7). For the pilot experiment Welch's t statistic was calculated for each protein for the difference between the TvO ratio and the technical replicate ratio for each patient.

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Equation 7 Welch's t statistic. X_1 = mean TvO or TvG ratio, X_2= mean TvT and OvO or TvT and GvG ratios, or for the pilot experiment the technical replicate ratio. σ^2 = variance, n=number (as defined in the text). Derived from reference[513].

$$\nu = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma^4}{n_1^2(n_1 - 1)} + \frac{\sigma^4}{n_2^2(n_2 - 1)}}$$

Equation 8 Calculation of the degrees of freedom for Welch's t test (from reference[514]).

The degrees of freedom were calculated for each protein using Equation 8 and t tables were then used to calculate p-values. All protein p-values were pooled and then false discovery rate (FDR)-corrected using the Benjamini-Yekuteili method[515]. Significance was defined as an FDR-corrected $p<0.05$.

2.11.4 Protein Identity Mapping

Protein databases use database-specific identifiers (IDs) to denote a single unique protein sequence. These can vary both between databases and between versions of the same database. To compare proteomics experiments using different databases for peptide spectrum matching, protein identifiers were mapped to Uniprot accession numbers (AC). International Protein Index (IPI) identifiers from proteins identified

in the PAcIFIC experiments were mapped to Uniprot Accessions using the web-based Uniprot mapping tool[516].

For unmapped IDs (~5% of the total) a manual search was performed and a Uniprot archive reference assigned to proteins not represented in the Uniprot database. When different IPI numbers corresponded to several isoforms of a protein, the spectral counts were summed for each experiment leaving a total count for all the proteins derived from one gene for each experiment. For the TMT experiments, all the peptides corresponding to various isoforms of a protein were summed under the parent protein denoted with one unique AC. This generated a list of unique Uniprot AC references with corresponding TMT and PAcIFIC quantitative data.

For ontology and network analysis, the Uniprot accession numbers were mapped to Entrez GeneIDs using reference tables downloaded from Ensembl, the National Centre for Biotechnology Information (NCBI) Entrez Gene service and the Human Genome Organisation Gene Nomenclature Committee (HGNC) web-servers and custom Perl and R scripts. For selected cases, manual searches were conducted using the Uniprot and Entrez websites[517].

2.11.5 Protein Subcellular Localisation

The subcellular localisation of identified proteins was determined from the UniprotKB database. A summary description of protein localisation was generated from the primary localisation; defined as the first localisation listed in the Uniprot database. Only limited conclusions could be drawn from this summary localisation as proteins are rarely confined to one subcellular compartment and not all identified proteins were annotated. Localisation options were restricted to; Cell membrane, Cytoplasm, Cytoplasmic Vesicle, Nucleus, Mitochondrion, Endoplasmic Reticulum (ER), Golgi, Secreted and unavailable. For ambiguous primary localisations, the further localisation data were reviewed and the most appropriate of the summary options selected.

2.11.6 Identifying Prior Evidence for Protein Expression in OAC and Other Cancers

A literature search was performed using the Pubmed database[518]. A literature base was obtained by searching all of the Pubmed records using the Medical Subject Headings (MeSH); oesophagus, oesophageal, OAC, Barrett's, esophageal, esophagus, EAC linked by an OR operator together with the MeSH; adenocarcinoma OR cancer. Dysregulated proteins were individually searched against this literature base and matches reviewed. Proteins were searched against published microarray studies of OAC[60,61,171,277] using the Oncomine database (v 4.4.3)[519] to determine if the genes were significantly dysregulated between OAC and normal oesophagus (FDR-corrected $p<0.05$). Proteins were also searched manually against the reported results from two existing proteomic studies using OAC tissue[447,449]. If no evidence for dysregulation in OAC was identified using this strategy a Pubmed literature search was performed to identify evidence of dysregulation in any other cancer.

2.11.7 Hierarchical Clustering of Expression Data

Unsupervised hierarchical clustering was performed using the hclust function in R (version 2.15.1). In this function, each sample is initially assigned to an independent cluster. An agglomerative method is then used to combine the pair of clusters with the shortest Euclidean distance until only one cluster remains[520]. Each new cluster was represented by a branch on the dendrogram and the distance between branches was proportional to the Euclidean distance between clusters.

2.11.8 Label-Free Quantitative Proteomics

For label-free proteomic analysis, the precursor acquisition independent from ion count (PAcIFIC) method was used with some modifications[404].

2.11.9 Sample Preparation

Tryptic peptides were generated from tissue samples using the modified FASP method (mFASP, see 2.9.5). Following overnight proteolysis with trypsin, the digestion buffer was evaporated by vacuum centrifugation (SpeedVac) and tryptic peptides resuspended in 200 µL dH_2O. This was repeated twice before the dried tryptic peptides from each tissue sample were finally resuspended in 5% (v/v)

acetonitrile, 0.1% (v/v) FA in dH_20 and desalted on a Macro SpinColumn C18 (Harvard Apparatus). Peptides were eluted and then evaporated to dryness by vacuum centrifugation. Dried peptides were resuspended in 5% (v/v) acetonitrile, 0.1% (v/v) FA in dH_2O to a concentration of 67 µg/ml based on the predicted protein input to the tryptic digests, calculated from the protein concentrations.

PAcIFIC LC-MS/MS was performed by Dr. Young-Ah Goo, (University of Washington, Seattle) on an LTQ-Orbitrap mass spectrometer (LTQ-Orbitrap, ThermoScientific) equipped with a UPLC system (nanoACQUITY, Waters).

2.11.10 Liquid Chromatography for PAcIFIC

Resuspended peptide samples were loaded by injection at 6 µL/min for 2.5 minutes onto a 200 µm internal diameter, 20 mm long trapping column (New Objective, Woburn, MA, USA) packed with 200 Å, 5 µm Magic C18 AQ (Michrom, Auburn, CA, USA). Trapped peptides were then eluted at 2.5 µL/min using a variable solvent gradient into a 200 µm internal diameter, 150 mm long analytical column (New Objective) packed with 100 Å, 5 µm Magic C18 AQ (Michrom). The gradient was created using a combination of 0.1% (v/v) FA in dH_2O (solvent A) and 0.1% (v/v) FA in acetonitrile (solvent B).

Each peptide sample was injected from the analytical column into a captive spray ionisation (CSI) source (Bruker, Billerica, MA, USA) and subjected to tandem mass spectrometry (MS/MS) with 34 repeat injections using the following variable gradient: 0 – 34 mins, 95% (A) and 5% (B); 34-37 mins at 58% (A) and 42% (B); 37-37.75 mins, 15% (A), 85% (B); 37.75-43 mins, 95% (A) and 5% (B).

2.11.11 PAcIFIC Mass Spectrometry

For each injection, a series of 30 sequential precursor ion windows incrementing in steps of 1.5m/z were selected. For each window, an isolation width of 2.5m/z was used and all precursor ions per window subjected to CID with collision energy set to 35% and the same fragment ion detection method used as for the TMT experiments. Precursor ions from the range 400 m/z to 1420 m/z were analysed across 34 sample injections to yield a complete PAcIFIC MS/MS analysis in just over 24 hours.

2.12 Data Analysis for PAcIFIC Experiments

2.12.1 Peptide Identification

For the analysis of PAcIFIC data, Thermo RAW spectral files for each sample injection were converted to mzXML files (34 files in total per sample) and peak lists were generated and searched with Sequest software[521] against the IPI human database (version 3.49, containing 74017 entries).

For database searching, oxidised methionines were set as variable modifications and carbamidomethylated cysteines were set as fixed modifications. The proteolytic enzyme was specified as trypsin with one missed cleavage permitted and the precursor mass tolerance was defined as 3.75 Da.

2.12.2 Protein Identification

Protein identifications were made using the Transproteomic pipeline (TPP) suite of tools[522]. Sequest results were converted to pepXML files and peptide identification probabilities computed with PeptideProphet[522]. For all individual sample searches (34 pepXML files), a threshold of peptide probability was identified to maintain the false discovery rate (FDR) <1%.

Protein identifications were then made from the pool of pepXML files for each patient sample using the ProteinProphet statistical model[522]. Only proteins with two unique peptides were accepted and when peptides matched to multiple members of a protein family, the protein family was summarised as a protein group with a single Uniprot identifier. The number of spectra identified per protein was automatically extracted from the pepXML files during analysis with the TPP. This spectral count was used as a measure of protein abundance.

2.13 Network Analysis

2.13.1 Mapping Identified Proteins to Functional Clusters

The Linghu functional linkage network was filtered at a likelihood ratio threshold of >5.0 and >20 (LR5+ and LR20+ respectively). The filtered networks were then clustered using the MCL algorithm[523] using sequential inflation values from 2.0 to 6.0. The total list of unique proteins identified in each experiment and the lists of

differentially regulated proteins were mapped to the unique list of genes in the filtered network with the matching genes retained. For each dysregulated gene list, a control list containing the same number of genes was generated by random selection from the total list of mapped genes. This process was repeated 10,000 times and the results concatenated into a control gene list file. Dysregulated and control lists were checked to ensure there were no genes unique to the list type. The differentially regulated and control gene lists were then mapped to the LR5+ and LR20+ clusters at each inflation value.

2.13.2 Identifying Significantly Enriched Clusters

The mapped clusters for dysregulated and control lists were compared and a FDR-corrected p-value calculated for the probability that the dysregulated gene mapping could have occurred through random chance (see Appendix B for R commands and custom Perl scripts). Clusters were considered significantly enriched if the FDR-corrected p-value was <0.1. Cluster gene membership and connectivity were then obtained from the original LR5+ or LR20+ clusters and networks and significantly enriched clusters visualised using Cytoscape[524].

2.13.3 Cluster Visualisation

For Cytoscape networks, all nodes were binned into 9 categories according to their TvO and TvG expression and the nodes filled according to the colour assigned to the appropriate bin. Networks were structured using the yFiles organic or Cytoscape edge-weighted spring-embedded strategy with manual manipulation to minimise overlapping nodes and enhance the clarity of presentation.

The biological process ontologies for each cluster were determined using the online tool; Database for Annotation, Visualisation and Integrated Discovery (DAVID, version 6.7)[525]. The ontologies were ranked by the significance of enrichment calculated using the DAVID online method and corrected for multiple hypothesis comparisons by the Benjamini-Hochberg method[526]. The top three processes with an FDR-corrected $p<0.05$ were retained to annotate the clusters. A summary describing the functional class of the genes was also produced from manual inspection of the gene membership of the cluster.

2.13.4 Pathway Mapping

To visualise the context of dysregulated genes within known biological pathways, the dysregulated genes (intra-experimental FDR-corrected $p<0.05$) were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG, version 66.1) online pathway database using the KEGGArray tool[527]. Genes were coloured within the pathway context according to either TvO or TvG expression using a 7-point colour scale.

Genes more than 2-fold upregulated in both TvO and TvG comparisons were also analysed using the online Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, version 9.05)[528]. A network was constructed using multiple edge-types derived from various curated sources including co-publication, co-expression and experimental demonstration of protein-protein interaction from various external databases.

To enhance connectivity between small clusters generated from the supplied genelist, nodes were inferred using the STRING algorithm which derives a probabilistic confidence score for the likelihood an inferred node is functionally linked to experimentally observed nodes. A minimum score of 0.99 was required so there was very high confidence in the functional association between inferred nodes and the experimental data. The inferred nodes could be used to strengthen the evidence for a molecular process or pathway within an incomplete dataset to provide a basis for hypothesis testing. For network visualization, unconnected nodes and peripheral nodes were removed from the network to generate one connected cluster, structured using a spring-embedded strategy with manual correction to enhance the clarity of presentation.

Chapter 3: Evaluation of PLK-1 as a Therapeutic Target in OAC

3.1 Chapter Aims

a) To evaluate the expression of PLK-1 in a panel of oesophageal cell lines and oesophageal tissue and correlate the findings with TP53 genotype.

b) To develop a method to accurately determine cytostatic from cytotoxic phenotypes in cell lines.

c) To assess the effect of PLK-1 siRNA and kinase inhibition on viability in oesophageal cell lines.

d) To evaluate strategies to enhance the therapeutic window between cancer cell lines and non-transformed cells.

3.2 Introduction

It is proposed that the DNA damage checkpoint at the G2/M transition may represent a fragile point for targeted therapy in oesophageal cancer (1.3.8). PLK-1 controls the exit from this checkpoint and selective inhibitors are in early phase clinical trials. In oesophageal cancer, only OSCC cells have been therapeutically targeted with PLK-1 inhibitors (1.5.12). The role of PLK-1inhibition and strategies to enhance the therapeutic window between oesophageal adenocarcinoma and non-transformed oesophageal cells will be examined in this chapter.

3.2.1 Cell Lines as Cancer Models

Immortal cell lines can be used to test the effect of inhibitors or genetic changes on viability, proliferation and other dynamic properties not assessable directly in tissue biopsies. By using cell lines derived from the tissues representing the stages of oesophageal carcinogenesis, it is hoped that cancer cell-specific phenotypes can be uncovered. The rationale being that the cell lines, although not isogenic, represent, in part at least, the epigenetic and transcriptional background of human oesophageal tissue. Changes only present in the cancer cell lines may represent therapeutic targets with intrinsic cancer-specificity and therefore a lower potential for systemic toxicity.

3.3 Results

3.3.1 Verification of a panel of oesophageal cell lines

STR genotyping was undertaken to verify the identity of the cell lines selected to represent the histological stages of oesophageal carcinogenesis (Table 3-1). The results demonstrate an exact match for the observed and predicted alleles for all the cell lines. In CP-A, OE33 and FLO-1 cells not all of the predicted alleles were confidently identified. No contradictory alleles were observed, however, supporting the verification of these cell lines. A comprehensive search of cell line STR genotype databases produced no matches for the HEEpiC cell genotype providing reassurance that the cells are primary cells rather than a cancer cell line. As a further verification step, the growth pattern and cell morphology observed by light

microscopy was confirmed to agree with images available through the European Collection of Cell Cultures[463].

The oesophageal cancer cell lines are all predicted to exhibit loss of p53 function due to TP53 mutation[85,471]. In contrast, both CP-A cells and HEEpiCs possess wild-type p53. In cells with wild-type p53, gamma and UV radiation induce p53 and downstream transcriptional target expression including MDM2, p21 (WAF1) and, in the case of irreparable DNA damage, effectors of the apoptotic machinery including BAX[529]. Wild-type p53 can be inactivated in cancer by MDM2 overexpression and therefore identification of the wild-type allele is not enough to confirm p53 function[530]. To demonstrate p53 functional status, the response to DNA damage was examined in CP-A and OE33 cell lines with predicted wild-type and mutant TP53 respectively. CP-A cells were initially exposed to both gamma and UV-C radiation at variable doses and then p53 and p21 expression examined after 24 hours by western blotting. CP-A cells exhibited the expected dose-dependent increase in both p53 and p21 expression in response to DNA damage from both UV-C and gamma irradiation (Figure 3-1). The highest UV dose produced a lower p21 response which may reflect secondary compensatory mechanisms activated in the 24 hours between exposure and measurement.

OE33 and CP-A cells were irradiated at a dose of 10 Gy and cells harvested for western blotting at serial intervals over 48 hours to avoid missing the induction of p53 target genes. There was a gradual and maintained induction of p53 expression from 2 hours until at least 48 hours after gamma irradiation in the CP-A cells (Figure 3-2). This was accompanied by an induction in MDM2, p21 and lastly BAX supporting the presence of functional p53 in this cell line (Figure 3-3). The time course of gene induction mirrors previous reports in MCF-7 breast cancer cells[531].

In contrast, the OE33 cell line exhibited no change in the expression of p53 or p53 transcriptional targets (Figure 3-2, Figure 3-3). This fits with the loss of p53 transcriptional activity due to a mutation in the DNA-binding domain (C135Y). Together these data support the expected p53 functional status of the oesophageal cells used in this study (Table 2-1).

Table 3-1 STR Genotype Profiling of Oesophageal Cells. ATCC – American Tissue Culture Collection, DSMZ -Deutsche Sammlung von Mikroorganismen und Zellkulturen, JCRB - Japanese Collection of Research Bioresources, RIKEN – Japanese Institute of Physical and Chemical Research

Sample	Allele at STR Locus									
	D21S11	THO1	TPOX	vWA	Amelogenin	CSF1PO	D16S539	D7S820	D13S317	D5S818
HEEpiC Observed	29,30	9.3	8, 11	14,16	X,Y	12	11,12	8,9	12,13	11,13
HEEpiC Predicted[474]	No Match Found on a Search of 2455 Cell Line STR Genotypes from ATCC, DSMZ, JCRB and RIKEN databases									
CP-A Observed	27,31.2	7,9.3	8,9	17	X,Y	12	12,13	7,8	14	11,12
CP-A Predicted[532]	-	7,9.3	8,9	17	X,Y	12	12,13	7,8	12,14	11,12
OE19 Observed	30	8,9	8	16,17,18	X	11,13	12,13	8	9,11	11,14
OE19 Predicted[85]	30	8,9	8	16,17,18	X	11,13	12,13	8	9,11	11,14
OE33 Observed	29,31.2	7	8	17	X	10,11	12	9,10	14	11
OE33 Predicted[85]	29,31.2	7,8	8,11	17	X	10,11	12	9,10	14	11
OACM5.1 C Observed	28,31	6,9.3	8	19,20	X	10,13	10,11	8	11,12	12
OACM5.1 C Predicted[85]	28,31	6,9.3	8	19,20	X	10,13	10,11	8	11,12	12
FLO-1 Observed	30,32.2	6	9,11	16		11	12,13	8	11	12,14
FLO-1 Predicted[85]	30,32.2	6	9,11	16	X	11	12,13	8	11	12,14
JH-Eso-Ad1 Observed	30	6,7	8,9	18,19	X	10	10,12	10,12	11	11
JH-Eso-Ad1 Predicted[85]	30	6,7	8,9	18,19	X	10	10,12	10,12	11	11
KYSE-30 Observed	28	9	8,9	16,18,19	X	10	10,12	11,11.3	9	11
KYSE-30 Predicted[474]	-	9	8,9	16,18,19	X	10	10,12	11,11.3	9	11

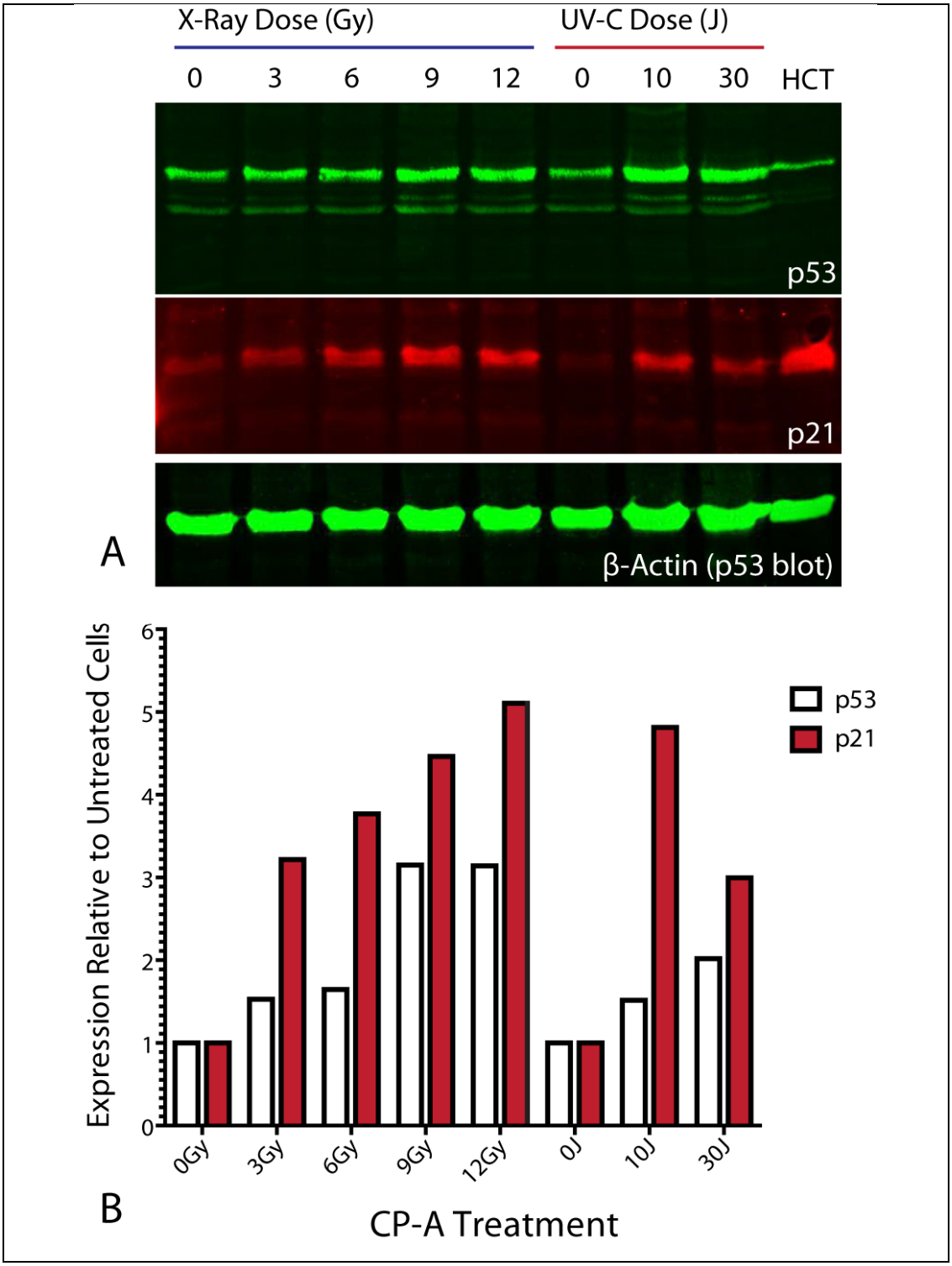

Figure 3-1 Response of CP-A Cells to DNA Damage. (A) CP-A cells were harvested at 24 hours after treatment, pellets lysed in urea buffer and 20 µg of protein (10 µg protein for HCT116 cells for p53 blot) used for SDS-PAGE (p53 - 12% agarose gel; p21 -15% agarose gel) and western blotting. After blocking in 5% Milk, PBS-T, membranes were probed with primary antibodies to p53 (DO-1) or p21 and then in both cases β-Actin. Secondary incubations were with Anti-mouse^{800} (p53, B-Actin) and Anti-mouse^{680} (p21). HCT – urea lysate from HCT116 cells, 24 hours after gamma irradiation (10 Gy). (B) Using Odyssey SA software, the band intensity (dominant band for p53) relative to β-actin from the same lane was used to quantify protein expression and then presented relative to the expression in untreated cells.

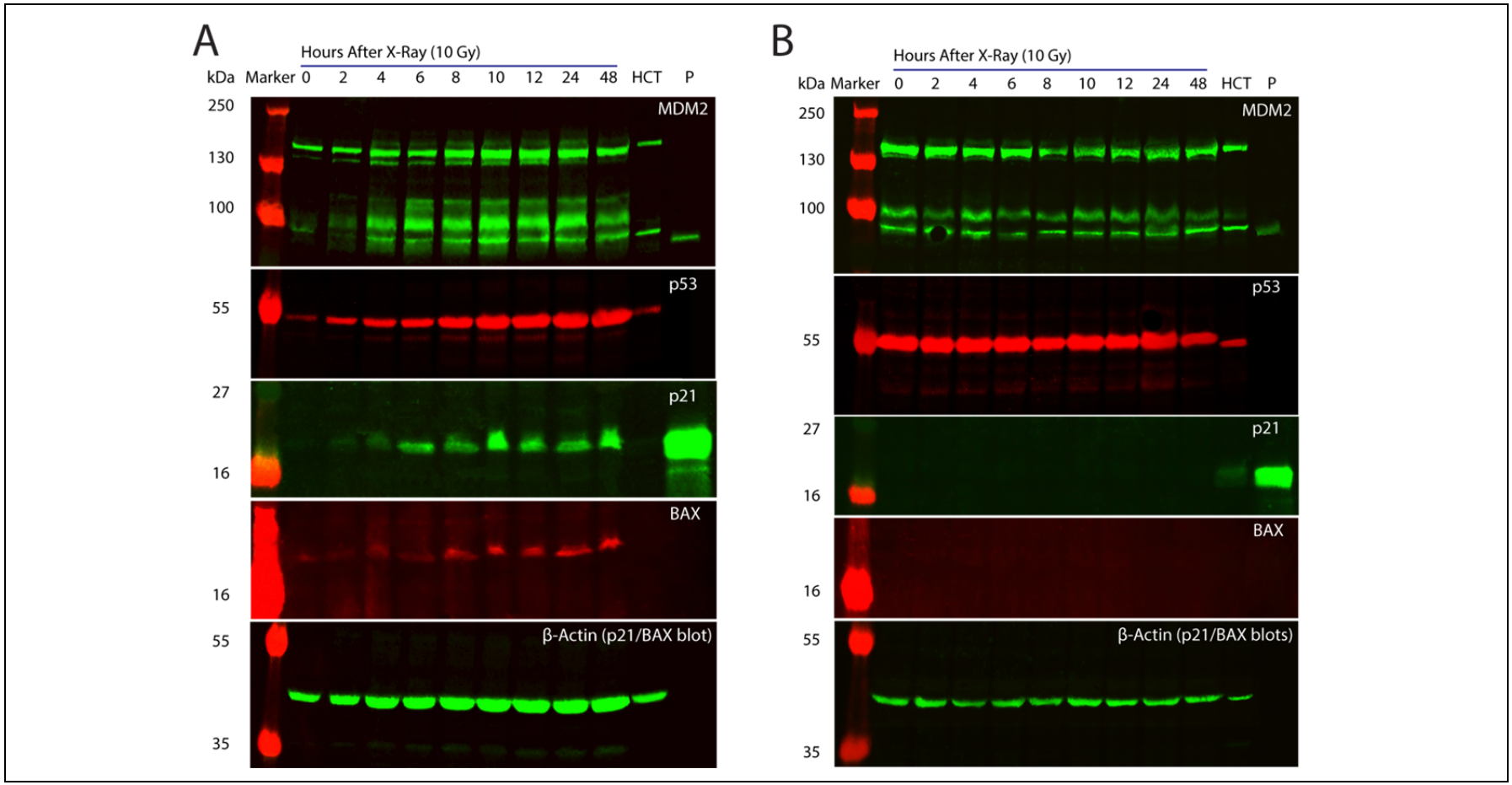

Figure 3-2 Comparison of the p53 response to DNA Damage in CP-A (A) and OE33 (B) Cell Lines. Cells were treated by gamma irradiation (10 Gy) and then harvested at fixed intervals. Pellets were lysed in urea buffer and 20 µg of protein (10 µg for HCT) used for SDS-PAGE (MDM2, p53 - 8% agarose gel; p21, BAX -15% agarose gel) and western blotting. HCT - urea lysate from HCT116 cells harvested 24 hours after gamma irradiation (10 Gy). P - Purified recombinant p21 and MDM2 protein. Membranes were probed sequentially with primary antibodies to MDM2 and p53 (CM-1) or p21 and BAX and then, in both cases, β-Actin. Secondary incubations were with Anti-mouse^{800} (MDM2, p21, β-Actin) and Anti-rabbit^{680} (p53, BAX).

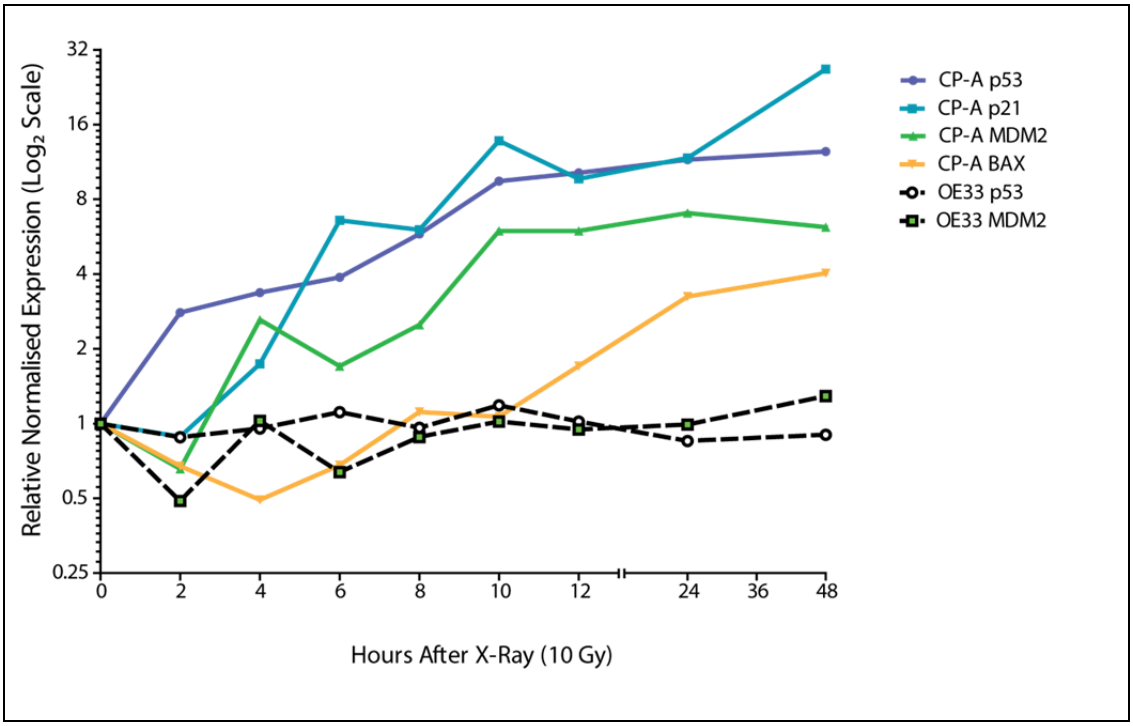

Figure 3-3 Quantitation of Protein Expression from Figure 3-2. The dominant band intensity was determined for each protein. For MDM2 expression in CP-A cells, the ~90 kDa band corresponding to the lower band from HCT116 cells and of similar size to the purified recombinant MDM2 protein was quantified. For p53 expression in CP-A cells, the dominant band corresponding to the p53 band in HCT116 cells was quantified. Expression was normalised to β-actin from the same lane and then relative to the protein expression in untreated cells.

3.3.2 Expression of PLK-1 in Oesophageal Cancer

The expression of PLK-1 was next investigated across the oesophageal cell line panel. PLK-1 expression was confirmed, as expected, to be higher in cells with mutant p53 (Figure 3-4)[220]. To confirm this was not simply an *in vitro* artefact, the expression of PLK-1 was also examined in oesophageal tissue representing the stages of Barrett's carcinogenesis. The specificity of the PLK-1 antibody used for this study (clone 35-206) was first determined by transfection of CP-A cells with siRNA to PLK-1 and quantitative western blotting.

Initial attempts at siRNA transfection in CP-A cells resulted in significant cytotoxicity, limiting the yield of cells and potentially leading to secondary, non-siRNA mediated effects on gene transcription. A comparison of transfection regimes was made with subjective visual assessment of cytotoxicity by light microscopy (Table 3-2). This revealed RNAiMAX produced the least cytotoxicity in both CP-A and OE19 cells and this reagent was used for all subsequent experiments.

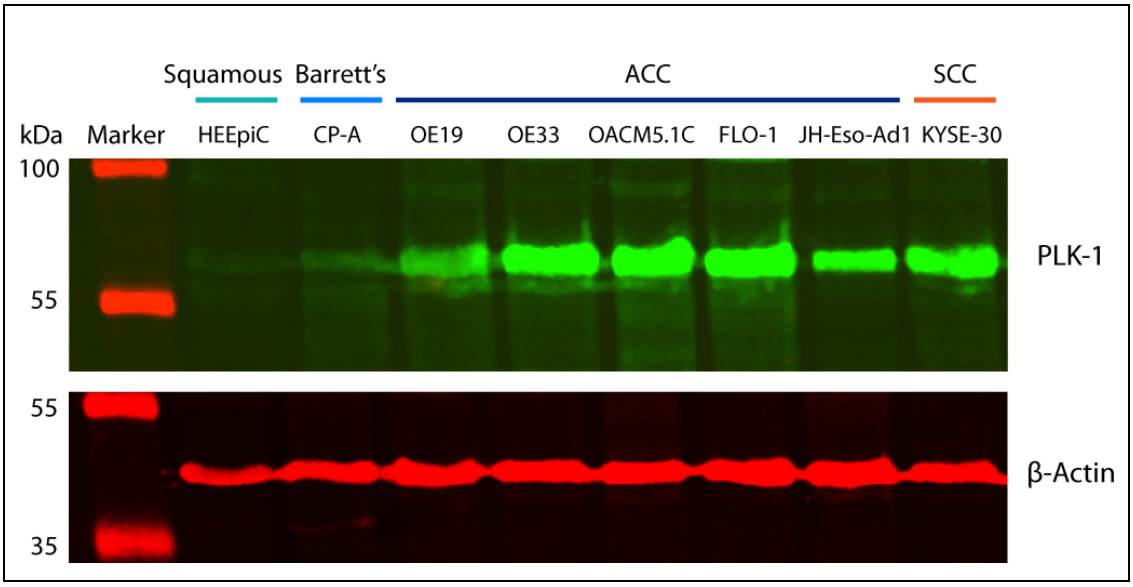

Figure 3-4 PLK-1 Expression in Oesophageal Cells. Cells were harvested from subconfluent plates and lysed in urea buffer before SDS-PAGE (20 µg protein per lane, 10% agarose gel) and western blotting. Membranes were sequentially incubated with antibodies to PLK-1, anti-mouse^{800}, β-Actin and anti-mouse^{680}.

Table 3-2 Comparison of Cytotoxicity from Transfection Reagents.

Cell Line	Visual Assessment of Cytotoxicity, 48 hours after media supplementation. (+++ = highly cytotoxic, - = no effect)				
	Opti-MEM I + Glutamax (20% v/v)	Dharmafect I (0.5% v/v)	Dharmafect 4 (0.5% v/v)	Lipofectamine 2000 (0.25% v/v)	RNAiMAX (0.325% v/v)
CPA	-	+++	+++	++	+
OE19	-	+	+	+/-	+/-

The specificity of the PLK-1 antibody (clone 35-206) was confirmed by the observed reduction in expression of the ~67 kDa band predicted to be PLK-1 after treatment with siRNA specific to PLK-1 (Figure 3-5). As a positive control for transfection, CP-A cells were also treated with siRNA to GAPDH and this reduced the expression of the single ~35 kDa band detected, confirming both the transfection conditions and the specificity of this antibody towards GAPDH (Figure 3-5).

There was also an apparent reduction in PLK-1 protein expression after transfection of non-targeting scrambled siRNA. Off-target effects of siRNA have been widely reported in the literature and can be mediated by partial complementation of

scrambled sequences[533,534]. Evaluation of PLK-1 mRNA expression was undertaken to examine this.

Relative changes in PLK-1 mRNA levels between conditions can be determined by qRT-PCR if the PCR amplification efficiency is known. The gradients of the linear regression lines from a cDNA dilution series were used to calculate the reaction efficiencies as; PLK-1 = 90%, AURKA = 91% and β-Actin = 91% (Figure 3-6). For subsequent gene expression assays using this method, a reaction efficiency of ~90% was assumed. As a further validation of the ΔΔCt method the ΔCt was found to be relatively constant for both PLK-1 and a second mitotic kinase; AURKA, across a wide range of input cDNA concentrations (Figure 3-6 (B)).

Quantitative real-time PCR was undertaken to confirm that the siRNA pools targeting PLK-1 and GAPDH specifically reduced PLK-1 and GAPDH mRNA levels and to investigate if non-targeting siRNA reduced PLK-1 mRNA expression. This demonstrated that both gene-specific siRNA pools significantly reduced expression but both the non-targeting, scrambled siRNA sequence and the siRNAs directed against GAPDH were associated with lower PLK-1 gene expression (Figure 3-7). This provides evidence for a non-specific effect of siRNA transfection on PLK-1 expression.

The relative expression of PLK-1 was next examined in tissue cores from an oesophageal tissue microarray. The same antibody confirmed to be specific for PLK-1 by western blot was used for IHC and, after optimisation of staining conditions, slides containing cores representing normal oesophagus (n=6), Barrett's (n=3) and OAC (n=12) were stained (Figure 3-8).

This revealed the majority of cells were negative for PLK-1 in normal squamous oesophagus. A few scattered PLK-1 positive cells were noted predominantly in the basal proliferative zone (Figure 3-8 (i)). Similarly in Barrett's crypts, a few isolated clusters of positive cells were noted but the majority were negative for PLK-1 (lower half of the core in Figure 3-8 (ii)). The OAC tissue, in contrast, revealed a higher frequency of PLK-1 positive cells. There was a more heterogenous staining pattern

across the OAC tissues with a range from ~10% to >50% of cells staining strongly positive (Figure 3-8 (iii), (iv)). Although the number of tissue samples assessed was relatively small, these data suggest that PLK-1 is over-expressed in OAC relative to surrounding normal tissues and that the oesophageal cell line panel may recapitulate the *in vivo* expression pattern.

Loss of p53 has been associated with upregulation of PLK-1 *in vitro* and this may explain the cell line and oesophageal tissue findings. Clarifying the relationship between p53 mutation and PLK-1 expression *in vivo* may allow those tumours with the highest PLK-1 expression, and potentially greatest sensitivity to PLK-1 inhibitors, to be identified using genetic testing.

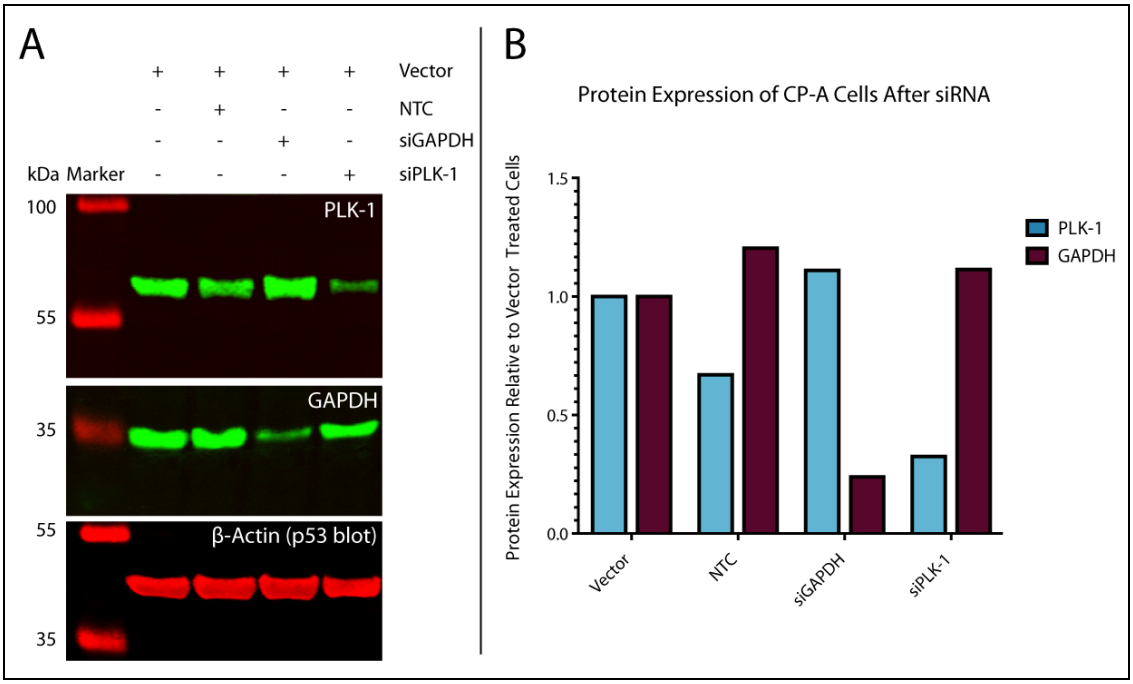

Figure 3-5 Protein Expression in CP-A cells after siRNA Transfection. (A) Cells were transfected with vector only, a scrambled non-targeting sequence (NTC), or siRNA targeting GAPDH or PLK-1 and harvested after 48 hrs. Pellets were lysed in Urea buffer before SDS-PAGE (20 µg protein per lane, 10% agarose gels) and western blotting. Membranes were sequentially incubated with antibodies to PLK-1 or GAPDH and then anti-mouse^{800} followed by β-Actin and anti-mouse^{680}. (B)Protein expression normalised to β-actin and relative to vector treated cells.

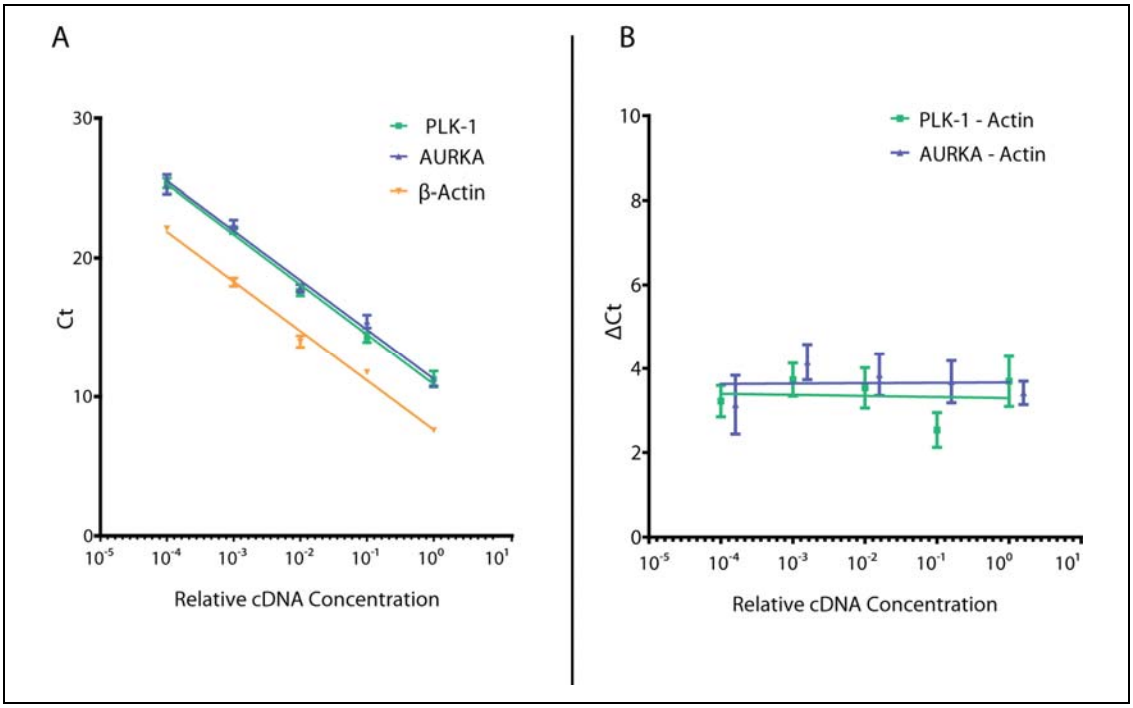

Figure 3-6 Evaluation of qRT-PCR Performance. (A) Efficiency of qRT- PCR was calculated using a dilution series of cDNA from a single stock. Points and error bars represent the Mean +/- SEM from 3-4 technical replicates. Linear regression after log_{10} transformation of cDNA concentrations revealed gradients and R^2 of; PLK-1 = -3.591, R^2 = 0.9954; AURKA =-3.556, R^2 = 0.9919; β-Actin = -3.565, R^2 = 0.9924. (B) Calculation of gene expression using ΔCt of PLK-1 or AURKA relative to β-Actin over a range of input cDNA concentrations. Best fit linear regression lines are illustrated.

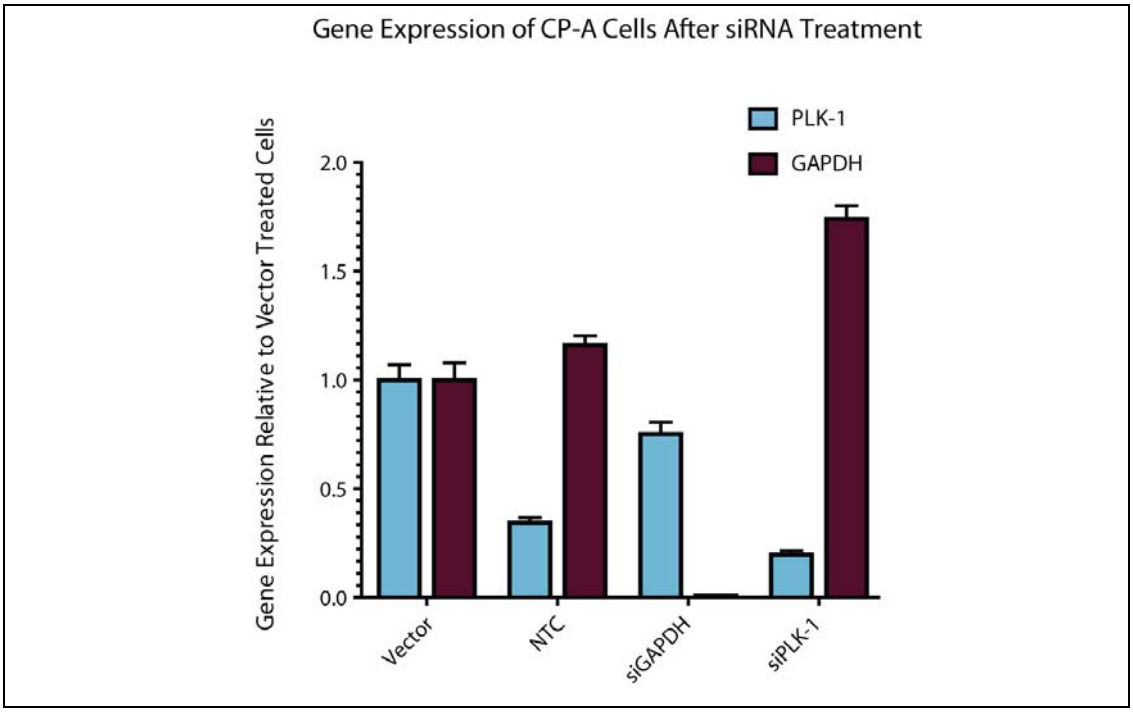

Figure 3-7 Effect of siRNA on Gene Expression in CP-A cells. RNA extracted from cell pellets treated identically to those from Figure 3-5 were used for qRT-PCR assays. Bars represent mean +/- SEM from technical quadruplicates.

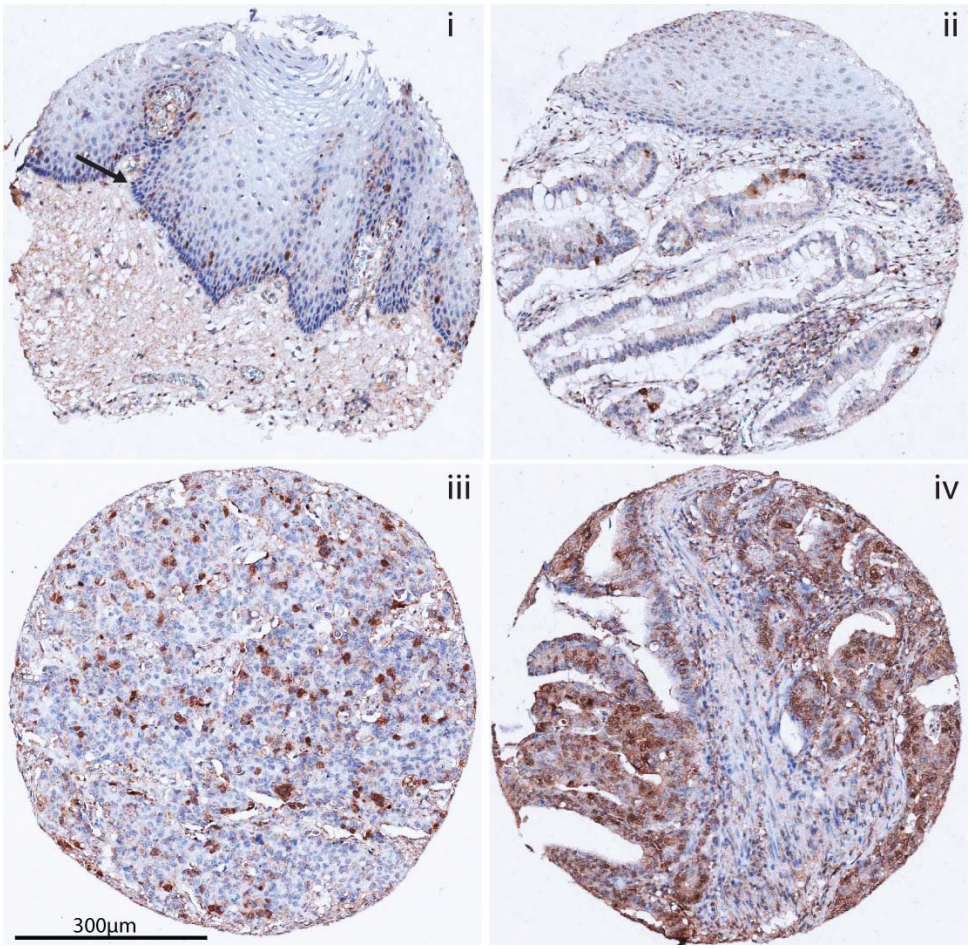

Figure 3-8 The Expression of PLK-1 in Oesophageal Tissue Cores .One hundred fold magnification images representative of; (i) normal squamous oesophagus - 6 cores; (ii) non-dysplastic Barrett's - 3 cores; (iii) + (iv) OAC - 12 cores. The arrow represents the squamous epithelial basal proliferative zone.

A tissue microarray created from a large cohort of oesophageal adenocarcinomas was used to determine the expression of PLK-1 in oesophageal tissues with wild-type and mutant p53. The TP53 genotype for these tumours was unknown. A method of inferring the genotype from the p53 immunohistochemical staining pattern has, however, been reported[193]. This relies on stabilisation of mutant p53 protein which is detected as nuclear over-expression in tissue sections[206]. The availability of highly sensitive and specific antibodies to p53 allows the detection of low level wild-type expression[535,536]. This allows TP53 deletion mutants to be detected due to the absence of protein expression in comparison to wild-type stromal expression. Co-staining with the p53 transcriptional targets p21 and MDM2 has been reported to improve the prediction of wild-type and null status in selected cases.

3.3.3 Determining TP53 genotype in OAC using IHC

The p53 staining pattern was determined in a tissue microarray consisting of triplicate cores from 356 oesophageal adenocarcinomas. Example staining patterns for tumours predicted to have p53 deletion mutations (null), p53 wild-type (wt) and mutant (mt) genotypes are demonstrated in Figure 3-9.

During the TMA construction, three cores from each tumour were cut and were embedded on separate TMAs. Sections from each of these arrays were stained and the core images distributed to two independent scorers (this author and Dr. Rudolf Nenutil, Pathologist, Brno, Czech Republic), blinded to the location of the replicate cores on the arrays. Although triplicate cores were cut from each tumour, not all cores contained diagnostic material and some cores were lost during the staining process, reducing the total number of scorable cores. The correlation between predicted genotypes from both scorers was assessed (Table 3-3).

Table 3-3 Inter-observer Correlation of Predicted TP53 Genotypes for OAC TMA cores

		Observer 2 (expert)		
		Wildtype	Mutant	Null
Observer 1 (non-expert)	Wildtype	106	30	8
	Mutant	49	501	0
	Null	0	0	186

The kappa statistic for the prediction of genotype across 880 scorable cores was 0.8194, indicating excellent inter-observer agreement. Disagreements generally arose from the classification of intermediate staining tumours as either wild-type or mutant and low staining tumours as either wild-type or null. Further staining with antibodies to p21 and MDM2 was not undertaken due to limitations of material but may have helped to resolve these classifications.

Total Allred histoscores for p53 were also compared for these cores. The inter-observer spearman correlation coefficient was 0.900 for individual cores and 0.8880 for the median score per tumour (both $p<0.0001$). This indicates p53 IHC is robust enough to be reliably applied by both expert histopathologists and non-expert scorers. When the p53 histoscores were plotted, three peaks in frequency were demonstrated corresponding to predicted mutant, wild-type and null (Figure 3-10A). Although inter-observer agreement was excellent, discordant genotype predictions within the three cores from the same tumour were observed. These are likely to be due to a combination of genetic and staining heterogeneity and both intra and inter-observer variability.

As the contribution of each of these factors was unknown, tumours were classified as having either confident or low-confidence genotypes. Tumours with confident p53 genotypes were defined as having unanimous predictions from both scorers across all cores (n=208). The remaining tumours were considered to have low-confidence genotypes (n=148). The median histoscores for tumours with confident predicted p53 genotypes formed more distinct groups corresponding to mutant, wild-type and null genotypes (Figure 3-10B).

The accuracy of genotype predictions was assessed using actual p53 genotypes from 22 tumours present on the array that had been subjected to whole-genome sequencing. When the IHC data for these tumours were examined, three groups of median histoscores corresponding to wild-type, mutant and null were evident with some overlap in the classifications (Figure 3-10C). The separation of histoscores became clearer although still not definitive when only those tumours with confident-predicted genotypes were considered. This may reflect subtle interpretation of the subcellular staining pattern by the observers during the determination of predicted p53 genotype. For example, moderate nuclear positive p53 staining was considered more likely to represent a p53 mutation than cytoplasmic staining. The accuracy of classifications using confident-predicted genotypes was assessed from the predicted and actual genotypes (Table 3-4).

Table 3-4 Comparison of Actual and Predicted Genotypes

		Confident Predicted Genotype		
		Wildtype	Mutant	Null
Actual	Wildtype	1	1	0
	Mutant	0	10	3

From these data the confident prediction of mutant p53 genotype is 100% sensitive and 93% specific for actual p53 mutation. Only limited conclusions can be drawn from this analysis as the frequency of wild-type tumours was low. Using confident predicted genotypes across the OCCAMS TMA, only 10.1% of tumours were predicted to have the wild-type allele with 20.2% predicted to have mutations leading to loss of p53 expression and the remainder predicted to have stabilising mutations.

3.3.4 Expression of PLK-1 in OAC according to TP53 genotype

The classification of tumours into predicted p53 wt, mt and null by IHC was considered acceptably accurate and the expression of PLK-1 was determined by IHC on the same array. Inter-observer variability was higher but still acceptable for PLK-1 scoring (Spearman correlation coefficient for median tumour PLK-1 histoscore = 0.5156; $p<0.0001$). In contrast to p53 histoscores, there was an approximately Gaussian distribution of PLK-1 histoscores (Figure 3-11A).

This pattern was observed because all tumours exhibited some PLK-1 staining and the main source of variation was in the frequency of positive cells. When median PLK-1 histoscores were compared between groups defined by predicted TP53 genotype, those tumours with predicted mutant p53 exhibited significantly higher PLK-1 expression (Figure 3-11B, $p=0.0448$).

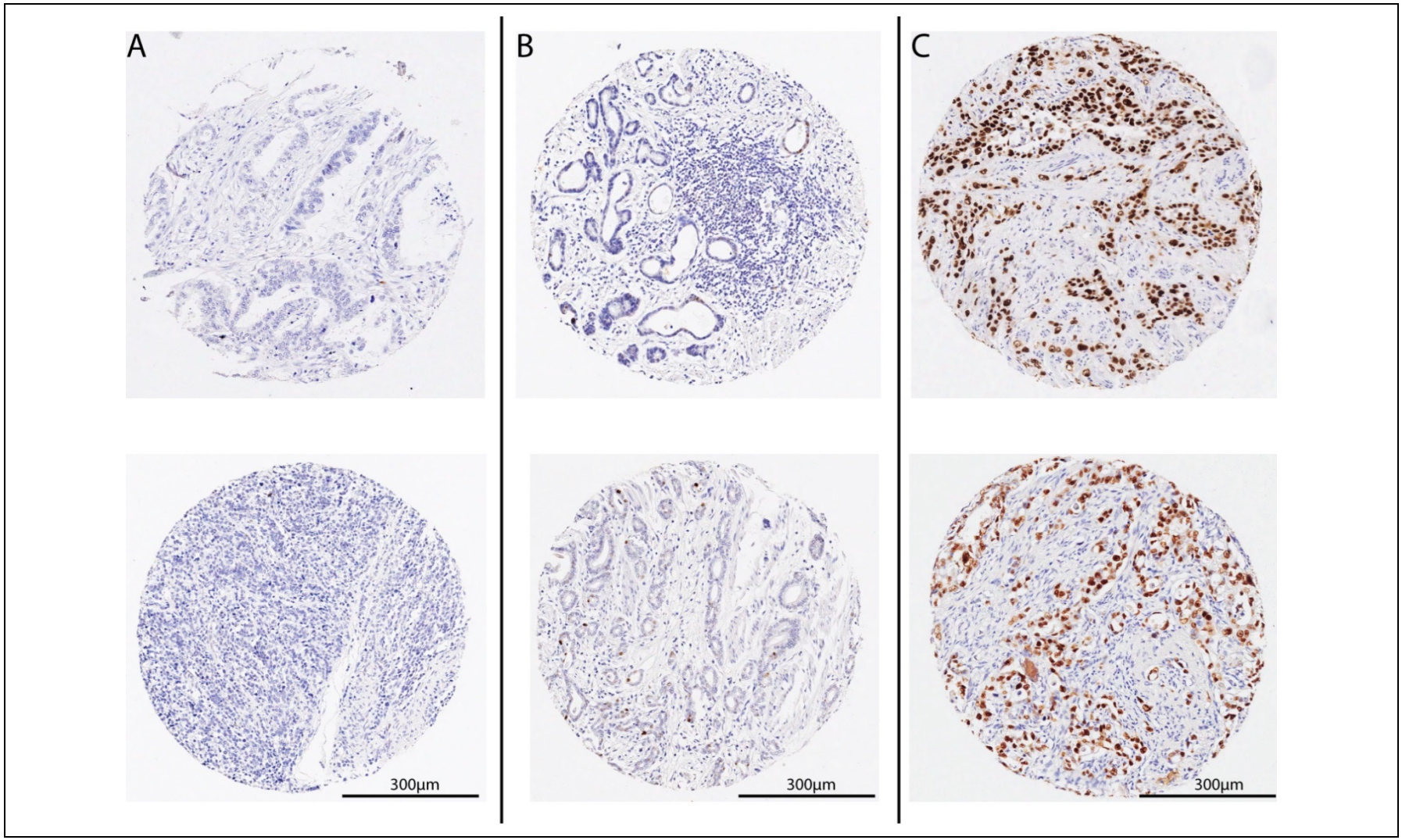

Figure 3-9 Representative Images of TMA Cores Stained for p53. Cores were predicted to have (A) p53 null, (B) p53 wt and (C) p53 mt genotypes.

Although only tumours with confident predicted TP53 genotypes were used in this comparison and the validity of this method is proposed, the p53 genotype of this cohort has not been definitively determined. No large cohort of TP53 genotyped OACs was available to assess PLK-1 expression. In an attempt to validate these findings and assess their applicability outside oesophageal cancer, the correlation of PLK-1 expression and TP53 status was tested in an external cohort of ovarian cancers with known TP53 genotypes.

PLK-1 expression also followed Gaussian distribution in ovarian tumours (Figure 3-11C). When median ovarian tumour PLK-1 histoscores were compared between groups defined by actual TP53 genotype, PLK-1 expression was again found to be significantly higher in tumours with mutant p53 (Figure 3-11D, $p=0.0279$).

The differential expression of PLK-1 in OAC compared to surrounding tissues may represent a unique dependency on PLK-1 for proliferation or cell survival. This could be tested by examining the viability of cells representing the stages of oesophageal carcinogenesis after reduction of PLK-1 levels by siRNA. To undertake this, a viability assay demonstrating a linear response with cell number was required.

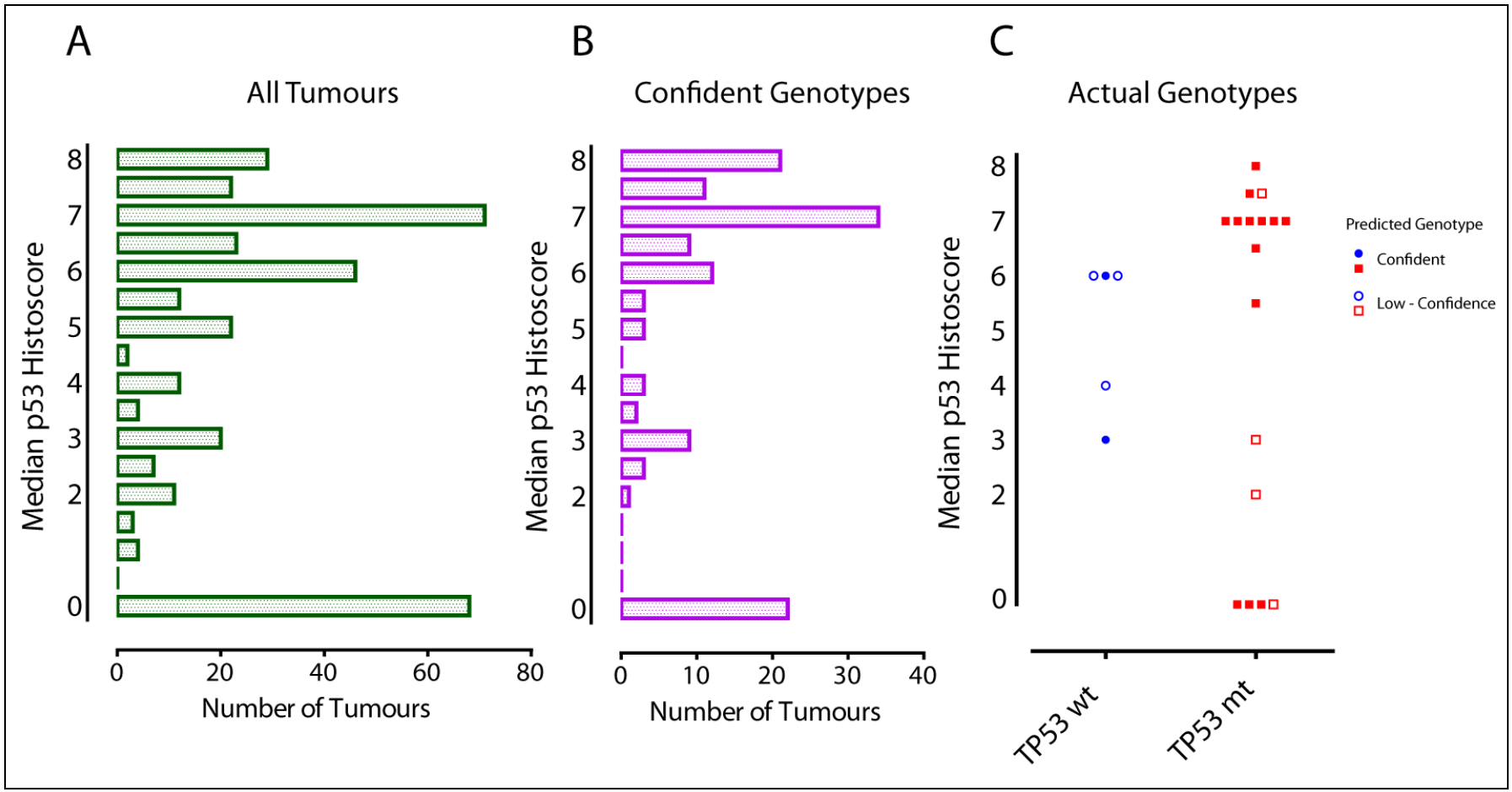

Figure 3-10 Distribution of OAC p53 histoscores from the OCCAMS TMA. The median p53 histoscore across all cores, from both scorers is plotted. (A) All tumours (n = 356). (B) Tumours with confident (unanimous) predicted genotypes. (C) The median histoscores of tumours with confirmed p53 genotypes from genome sequencing. Those tumours with confident predicted (n = 15) and low-confidence predicted genotypes from IHC (n = 7) are highlighted.

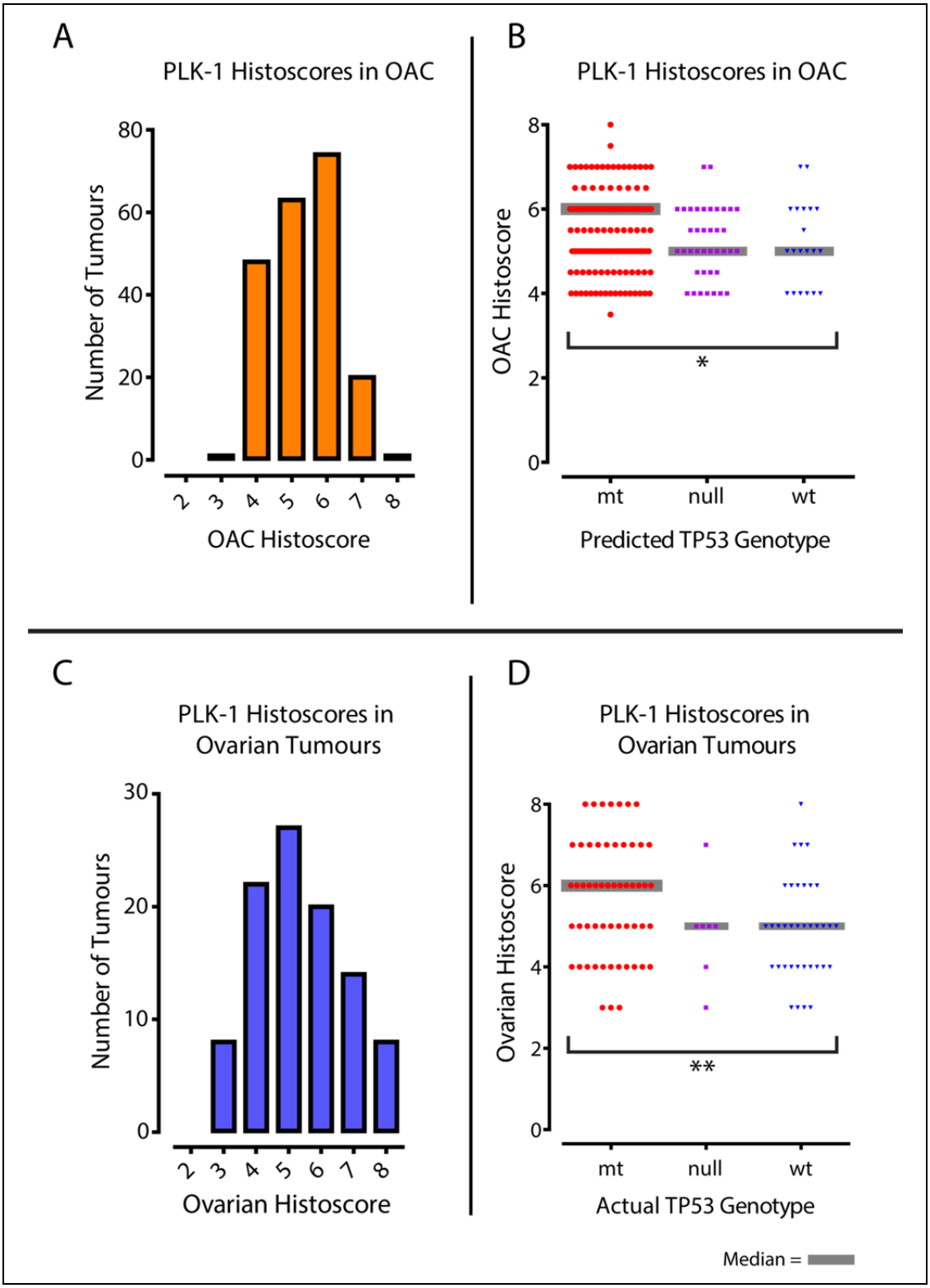

*Figure 3-11 PLK-1 Expression in OAC and Ovarian Tumours. (A) Median PLK-1 tumour histoscore from two independent observers (B) Scatterplot of median PLK-1 histoscores in tumours classified according to predicted TP53 genotype (confident predictions only n=208). (C) Median PLK-1 histoscore in ovarian tumours with known TP53 genotypes (n=99). (D) Scatterplot of median PLK-1 histoscores in tumours classified according to actual TP53 genotype. (B,D) Grey bars represent median value, * p=0.0448, **p=0.0279.*

3.3.5 Identification of an Optimal Viability Assay

The CellTiter-Glo assay was initially tested and provided a linear response with increasing cell number (r^2>0.99, Figure 3-12). Assays relying on luminescence detection are often performed in white opaque plates to enhance signal detection. Unfortunately these plates do not allow periodic monitoring of cell growth in culture. To compare the effect of plate type on CellTiter-Glo readings, reaction mixtures from the CellTiter-Glo assay were transferred from the clear-bottomed, black-walled wells of a 96-well tissue culture plate to the white-walled, white-bottomed wells of a 96-well plate and luminescence re-measured. An approximately 10-fold increase in signal intensity was observed when the white plates were used (data not shown). When the relative signal intensities were compared, however, no change in the response of the assay was detectable between plates (Figure 3-12 A). This indicates reliable performance with oesophageal cells in a 96-well format.

The disadvantage of the CellTiter-Glo assay is that the reagent lyses cells and therefore only provides endpoint measurements. Preliminary experiments demonstrated variability in seeding density across cell lines due to difficulties in generating a uniform single cell suspension. By measuring final cell number only, variations in initial seeding density are not controlled. Determining the final cell number relative to the initial cell number was thought to be relevant as treatments that reduce the cell number below the initial seeded number were considered "truly" cytotoxic. In contrast, treatments that produce a reduction in cell number compared to untreated cells but allow an increase in cell number from baseline were considered either partially cytotoxic or cytostatic.

The AlamarBlue reagent is reported to be water-soluble and non-toxic in both the native and the reduced fluorescent form[486]. This allows the reagent to be washed off by serial dilution in fresh tissue culture medium and allowed an initial measurement of cell number after cells have adhered to tissue culture plates and before treatment. After the treatment period a final cell number measurement can be taken, allowing the effect of the treatment to be normalised for initial cell number on a well by well basis and revealing both growth inhibitory and truly cytotoxic phenotypes.

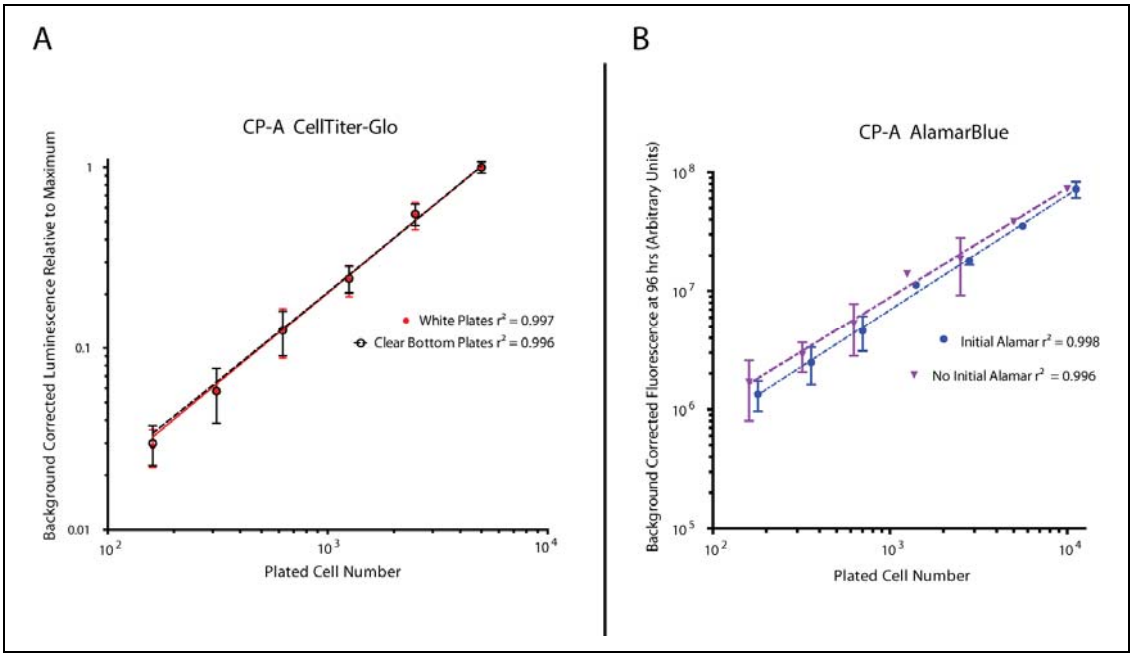

Figure 3-12 Comparison of CellTiter-Glo and AlamarBlue. CP-A cells were seeded into clear-bottomed black-walled 96 well plates at a range of densities and incubated for 24 hours. (A) Luminescence was measured (arbitrary units) after addition of CellTiter-Glo reagent. Reaction mixtures were transferred to white 96-well plates and luminescence measured again. Background corrected luminescence was normalised to the maximum value. (B) For the Initial Alamar experiment, an AlamarBlue measurement was taken 24 hours after seeding, the cells washed and then incubated in fresh medium for a further 72 hours followed by a final AlamarBlue measurement. Alternatively a final AlamarBlue measurement was taken 96 hours after seeding without an initial Alamar measurement or a change of medium. Background-corrected fluorescence from the final AlamarBlue measurement is plotted. (A,B) Points represent trimmed means +/- SD from 8 technical replicates. Linear regression lines are displayed. (B) The gradients of best fit lines were not significantly different (ANOVA, p=0.8847).

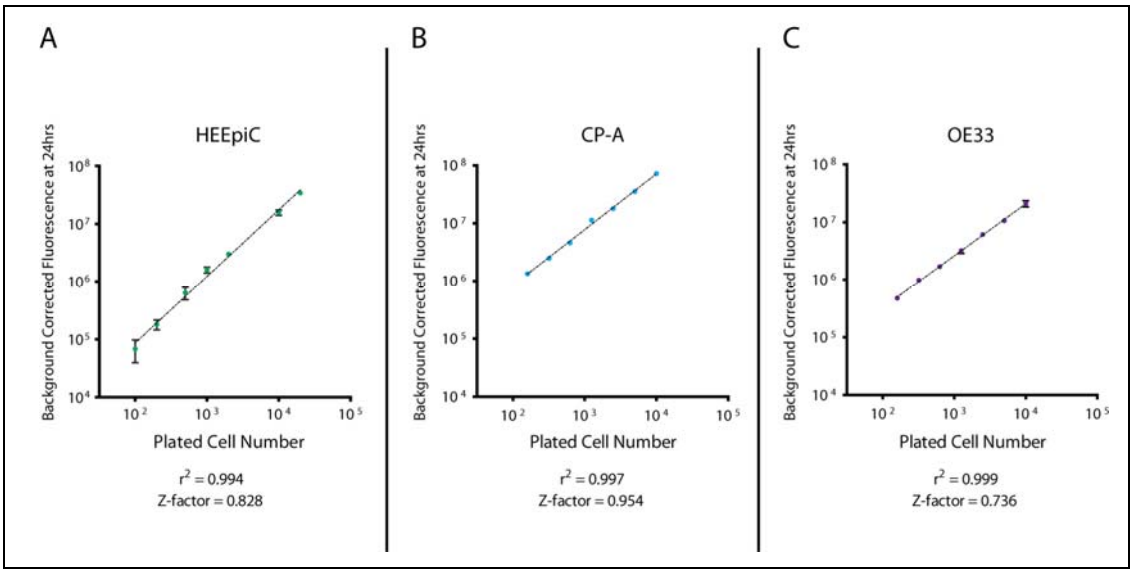

Figure 3-13 AlamarBlue Response in Oesophageal Cells. Cells were plated in poly-L-Lysine coated clear-bottomed, black-walled, 96-well plates and allowed to adhere for 24 hours. An AlamarBlue measurement was taken and background-corrected fluorescence calculated. Best-fit linear regression lines are shown. Points represent trimmed mean +/- SEM of 8 technical replicates.

Using CP-A cells, AlamarBlue was found to produce a linear response with increasing cell number, similar to CellTiter-Glo (Figure 3-13B). Importantly, an initial measurement of cell number followed by reagent wash-off did not significantly affect the final viability measurement compared to untreated cells (Figure 3-12 B). A linear response with increasing cell number was also demonstrated in primary oesophageal cells (Figure 3-13 A), and OE33 cells (Figure 3-13 C) with a low coefficient of variation and excellent assay performance (Z-factor 0.736-0.954). As a further control, wells treated with a non-ionic surfactant to induce cell death (5% v/v final concentration Triton-X-100) exhibited minimal AlamarBlue-induced fluorescence above background (data not shown). This confirms AlamarBlue allows specific and accurate measurement of viable cell number and this method was employed for subsequent viability assays.

3.3.6 The Effect of PLK-1 siRNA on Viability

The effect of selective PLK-1 gene knockdown on oesophageal cell viability was examined using the AlamarBlue method to enable a comparison of final and initial cell number. The optimised siRNA transfection regimen expected to significantly reduce PLK-1 expression at the mRNA and protein level was used (Figure 3-5, Figure 3-7). Knock-down of PLK-1 significantly reduced cell number compared to transfection vector and scrambled siRNA treated CP-A, OE19 and KYSE-30 cells (Figure 3-14). When the same transfection protocol was applied with HEEpiC and OE33 cells, the transfection reagent induced significant cytotoxicity so that the effect of the siRNA could not be confidently determined (Figure 3-15 A). Attempts with other lipid-based transfection reagents produced similar results.

Cell-permeant siRNA (Accell®, Dharmacon) does not require lipid-mediated transfection and early reports of its use have suggested high transfection efficiency in primary cells, with reduced cytotoxicity compared to other transfection regimes[537]. Accell siRNA uptake efficiency was assessed by transfection of FAM-labelled scrambled siRNA with blue and white light bright-field images captured at 48 hours. This demonstrated uptake of the fluorophore in the majority of cells, suggesting the transfection protocol was effective (Figure 3-15 C,D).

Accell siRNA was then transfected into OE33 and HEEpiC cells and significantly reduced cytotoxicity compared to RNAiMAX was observed (Figure 3-15 B), although there was still a significant reduction in cell number compared to untreated controls. When a pool of siRNAs to PLK-1 was transfected, viability was further reduced compared to non-targeting scrambled sequence siRNA and siRNA to GAPDH in both HEEpiC and OE33 cells. In HEEpiC cells, siRNA to GAPDH also reduced viable cell number compared to a non-targeting control. This may be a consequence of the sensitivity of HEEpiC cells to transfection, dependency on GAPDH for proliferation or off-target effects of the siRNA.

From these data, transfection of siRNA to PLK-1 reduced viable cell number in all 5 cell types tested. The magnitude of reduction with respect to baseline cell number was, however, variable.

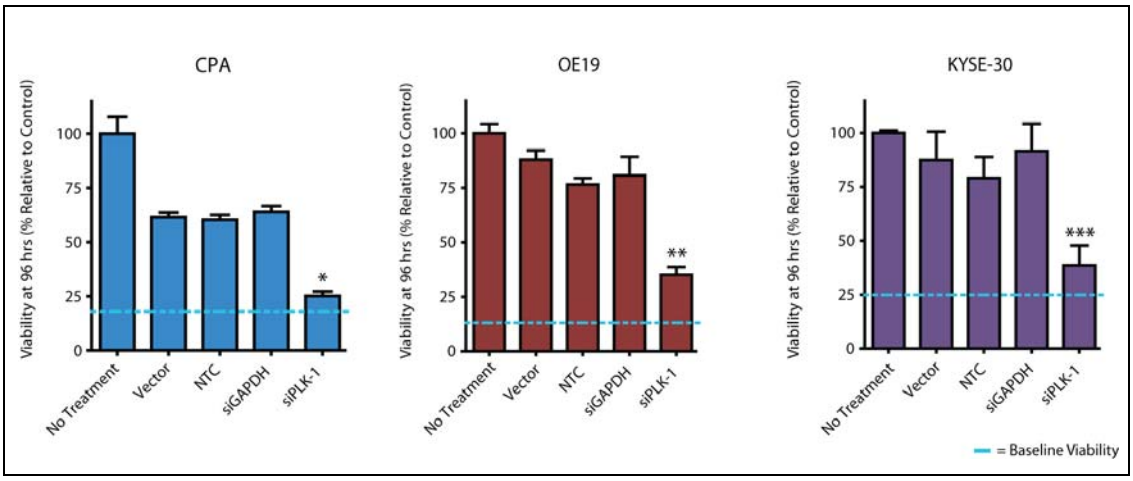

*Figure 3-14 The Effect of siRNA to PLK-1 on Oesophageal Cell Viability. After a baseline viability measurement using AlamarBlue, cells were treated with unsupplemented complete medium (No Treatment), RNAiMAX (Vector), scrambled non-targeting siRNA (NTC), or siRNA targeting GAPDH or PLK-1. Viability was assessed at 96 hours post-transfection and was normalised to No Treatment values. Bars represent mean +/- SEM of 6-12 technical replicates. The baseline viable cell number relative to the final viable cell number is presented for comparison. *, **, *** p<0.001 (unpaired, two tailed student's t-test), compared to vector treated cells.*

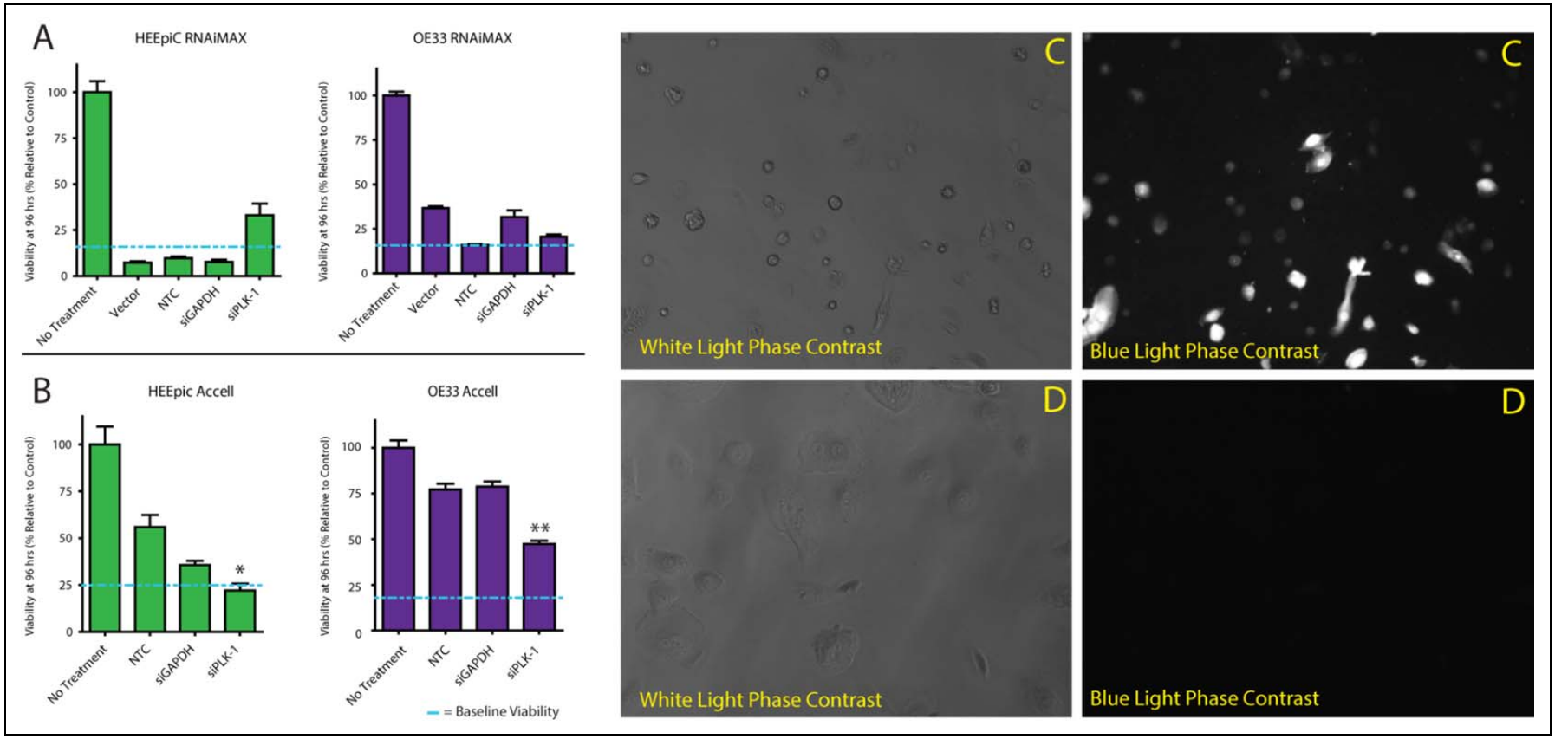

*Figure 3-15 (A, B) Comparison of the Effect of RNAiMAX (A) and Accell siRNA (B) on Viability at 96 hours. The experimental protocol was identical to Figure 3-14. NTC = non-targeting control siRNA. Bars represent mean +/- SEM of 6-12 technical replicates. Baseline viability is presented for comparison. *siPLK-1 vs NTC, p=0.0011, **siPLK-1 vs. NTC p<0.001, (both unpaired, two-tailed student's t-test). (C, D) Assessment of Accell transfection efficiency. HEEpiC cells were transfected with FAM-labelled (C) or unlabelled (D) non-targeting scrambled sequence siRNA. After 48 hours the medium was exchanged and images acquired using a digital camera and an inverted phase-contrast microscope with either white light (left images) or blue/UV light illumination at low power (C, 10x objective) or high power (D, 40x objective).*

3.3.7 Combining Viability and Gene Expression Measurements

To correlate the degree of PLK-1 suppression with viability, the In-Cell Western method was applied. In this technique, cells are fixed and permeabilised after siRNA treatment and viability measurement and then quantitative immunofluorescence performed to measure protein expression. The permeabilisation stage of the In-Cell Western method required several washes with detergent containing buffer. Despite coating the 96 well plates with poly-L-Lysine, most OE33 cells detached during this process and the In-Cell Western method was deemed unfeasible for this cell type.

To test the robustness of this method, protein expression was measured for several proteins by measurement of fluorescence at 800 nm from conjugated secondary antibodies. The ratio of fluorescence at 800 nm to fluorescence at 700 nm from the DNA stain, DRAQ5 was calculated for a range of cell densities. It was assumed that for most genes, protein expression should remain constant relative to cell number across a range of subconfluent cell densities in 96 well plates. Changes in viable cell number should not therefore change relative protein expression. For comparison of viable cell number measurements, fluorescence at 800 nm was also normalised to relative viability assessed by AlamarBlue assay prior to In-Cell Western.

Data were obtained for both CP-A and OE19 cells with minimal background fluorescence detectable in the secondary only and no-cell controls (Figure 3-16 A). Despite the favourable signal to noise ratio for fluorescence detection, there was significant variability in the 800/700 ratios across the cell densities (Figure 3-16 B). No pattern was observed with increasing cell number for any of the primary antibodies. The CP-A cells proved more reliable than the OE19 cells, potentially due to the flat epithelial growth pattern and resistance to detachment.

The variability in response was also evident when AlamarBlue fluorescence was used to normalise for cell number confirming that DRAQ5 fluorescence was not the only source of variation. The difference in variability between primary antibodies suggests either the target proteins exhibit highly variable expression pattern or the variability is due to the indirect method of protein detection. To confirm reduced protein expression after siRNA treatment, the In-Cell Western technique would need

to demonstrate accurate and sensitive measurement across oesophageal cell lines. These data do not support this. To measure viability and confirm gene knockdown after siRNA treatment, parallel experiments in 6-well plates with identical conditions were therefore performed.

Due to the effect of PLK-1 gene knock down on cell number, there were insufficient cells remaining 96 hours after treatment to allow RNA and protein expression analysis. To confirm sufficient gene knockdown, protein and RNA expression were therefore assessed 48 hours after treatment.

Using the optimised cell-specific siRNA conditions, PLK-1 gene expression could be effectively suppressed in all cell types and a corresponding significant reduction in viable cell number was observed after 96 hours treatment (Figure 3-17). A reduction in PLK-1 mRNA expression but not viability was observed with non-targeting siRNA transfection in CP-A, OE19 and KYSE-30 cells.

These data support the critical role of PLK-1 in oesophageal cell proliferation. PLK-1 has been reported to mediate cell cycle effects through both protein-protein interactions and substrate phosphorylation[232,538]. A reduction in protein expression would affect both mechanisms. The role of substrate phosphorylation can be specifically investigated using inhibitors of PLK-1 kinase activity.

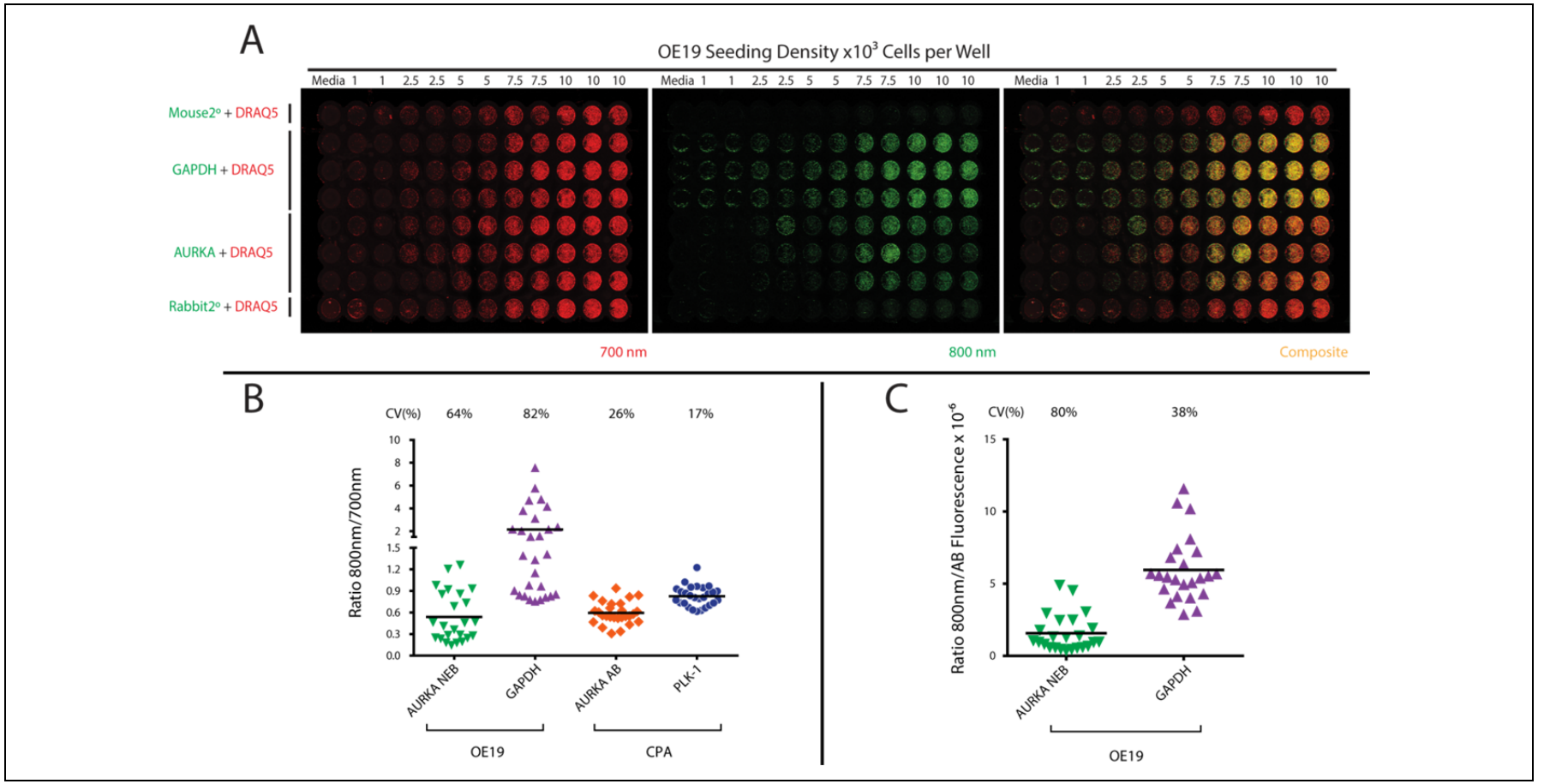

Figure 3-16 Evaluation of the In-Cell Western Method to combine viability and protein expression measurement. Cells were seeded at a range of densities including a no cell control (Media) by serial dilution from the same cell solution and grown in standard media for 48 hours. AlamarBlue and then In-Cell Western assays were performed. (A) Licor scan images obtained after In-Cell Western of OE19 cells using DRAQ5 and antibodies to GAPDH, AURKA (NEB, #4718) or secondary antibody only. (B) Scatterplots of 800/700 nm ratios for OE19 and CPA cells across the range of densities after In-Cell Western with antibodies to AURKA NEB (#4718), AURKA AB (ab52973), GAPDH or PLK-1. Bars represent medians with coefficient of variation (CV) presented as a percentage of the median. (C) Scatterplots of 800 nm AlamarBlue Fluorescence after In-Cell Western in OE19 cells using the same conditions as (B).

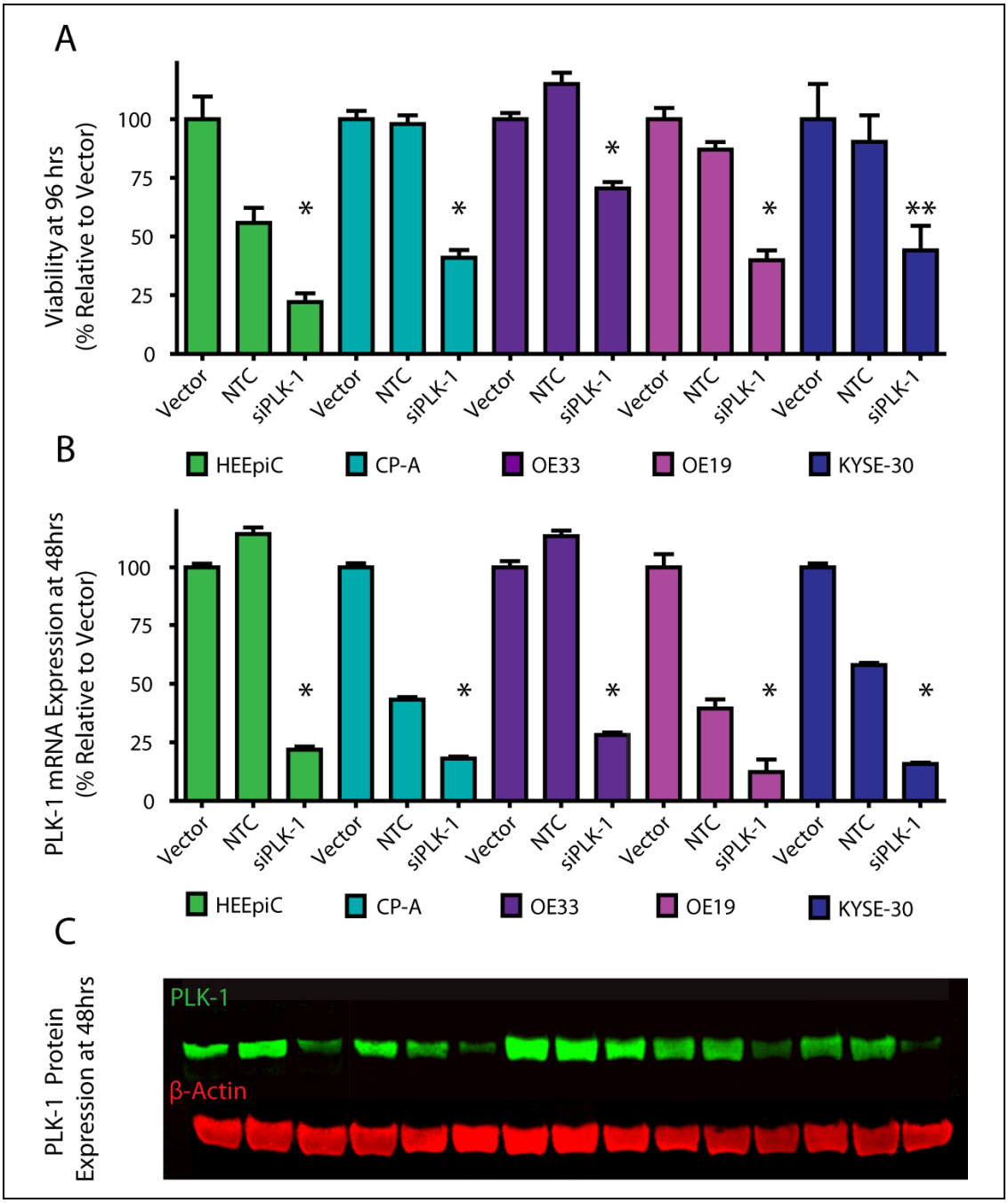

*Figure 3-17 The Effect of siRNA to PLK-1 on Oesophageal Cell Viability. (A)Cells were treated in 96-well plates. Vector = RNAiMAX; CP-A, OE19, KYSE-30 or complete medium; HEEpiC, OE33. NTC = Non-targeting scrambled RNA (Dharmacon; CP-A, OE19, KYSE-30 or Accell; HEEpiC, OE33). siPLK-1 = a pool of 4 siRNAs targeting PLK-1 (Dharmacon; CP-A, OE19, KYSE-30 or Accell; HEEpiC, OE33). Viability was assessed using the AlamarBlue method at 96 hours post-transfection. Bars represent trimmed mean of ≥ 8 technical replicates +/- SEM. (B,C) Cells were treated in 6 cm plates using identical conditions to (A), harvested for RNA (B) or protein (C) extraction at 48 hours post-transfection. (B) RNA expression was determined by qRT-PCR. Bars represent mean +/- SEM of 4 technical replicates. (C) Lysates (30 µg protein per lane) were resolved by SDS-PAGE (8% gels) and subject to quantitative western blotting by sequential incubation with primary antibodies to PLK-1, Anti-mouse^{800}, β-Actin and Anti-Mouse^{680} antibodies (Odyssey SA, Licor Biosciences). The image is a cropped membrane scan (100 µM resolution) from the 800 nm (green) and 700 nm (red) channels. *p<0.001, **p=0.012 compared to vector treated.*

3.3.8 Evaluation of PLK-1 Inhibitors

The first in class PLK-1 inhibitor, BI2536, mimics the response of cells to PLK-1 RNA depletion and is a potent and specific inhibitor of polo-like kinases[290]. For this study, a novel PLK-1 inhibitor was also obtained with a shorter half-life and a greater specificity for PLK-1 over other polo-like kinases (PLK-1A). This compound is derived from a pyrimido-diazepinone scaffold and, similar to BI2536, competes with ATP for the ATP-binding domain of PLK-1[539]. This blocks the kinase activity but not protein interactions with the polo-box domain. The PLK-1A compound is reported to have greater selectivity for PLK-1 over other polo-like kinases compared to BI2536 (Table 3-5)[540]. Human pharmacokinetic data are not yet available for PLK-1A as the compound has not yet been tested in phase I trials, but the half-life ($t_{1/2}$) in rats is reported as 0.3 hours. A study using animal pharmacokinetic data from 30 compounds and comparing various prediction models found human drug half-life could be predicted, with an average 1.79 fold error, from the function;

$$Log_{10}Human\ t_{1/2} = \left(0.906 \times Log_{10}Rat\ t_{1/2}\right) + 0.723$$

Equation 9 Prediction of Human Drug Half-life from Rat Data (from Reference[541])

Using this function, an appropriate drug dosing regimen could be predicted in 83% of cases. Applying this equation, the predicted human half-life would be 1.8 hours (estimated error range 1 – 3.2 hours). A comparative study of platinum-based compound plasma half-lives reported a 1.25-2.9 fold increase in half-life in humans compared to mice[542]. Directly translating these data and assuming drug half-lives in mice and rats are equivalent, the predicted human plasma half-life after intravenous infusion of PLK-1A would be between 0.4-0.9 hours. In either case, BI2536 would be predicted to have a roughly 10 fold greater half-life *in vivo*.

Table 3-5 Properties of PLK-1 Inhibitors – Adapted from Reference[540]

Compound	IC_{50} (nM) for Polo-like Kinases			Plasma Terminal Elimination Half-life (hr)
	Kinase			
	PLK-1	PLK-2	PLK-3	
PLK-1A	14 (1x)	231 (17x)	206 (15x)	Rat = 0.3
BI2536	7 (1x)	20 (3x)	11 (1.5x)	Human = 20-30hrs[315]

To confirm the inhibitors produced the expected phenotype, cells were treated for 24 hours and cell cycle changes profiled by flow cytometry after propidium-iodide staining[543]. Both the PLK-1 inhibitors produced an accumulation of 4N cells consistent with activation of the G2/M checkpoint (Figure 3-18). This was noted in non-transformed and cancer cell lines and suggests that the tested compounds are *bona fide* PLK-1 inhibitors.

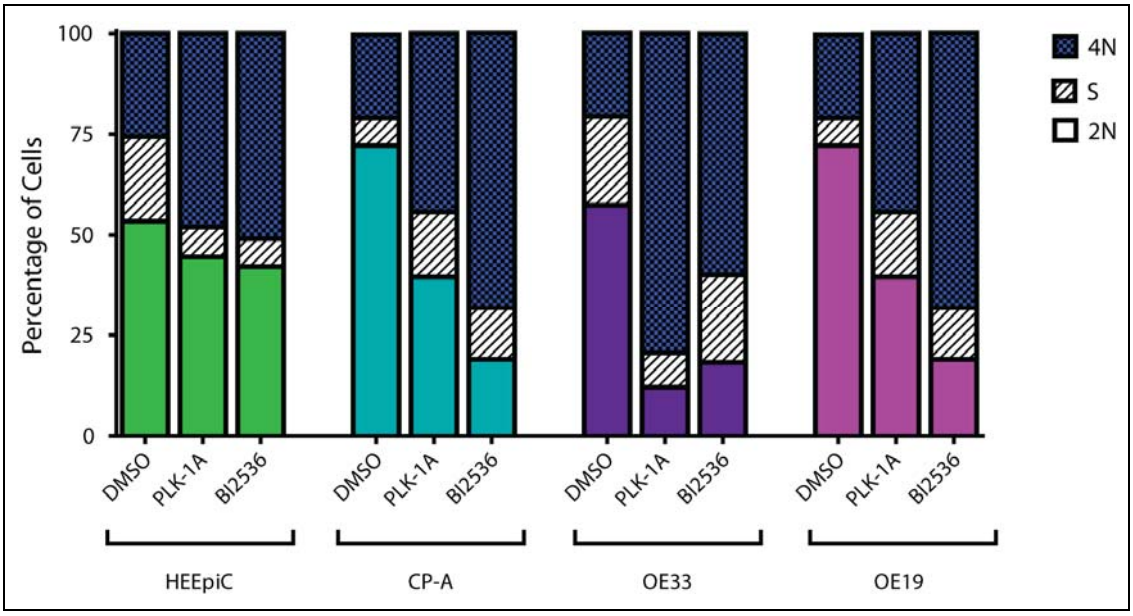

Figure 3-18 The Effect of PLK-1 Inhibitors on Oesophageal Cell Cycle Stage. Cells were synchronised by serum or supplement starvation and then released into complete medium with either DMSO, PLK-1A (1 μM) or BI2536 (50 nM). Cell cyle profiles were then determined after 24 hours by propidium iodide staining and flow cytometry. Results are representative of two biological replicates.

The effects of the PLK-1 inhibitors on viability were next examined in oesophageal cells using the AlamarBlue method. This revealed dose-dependent reduction of viable cell number for both PLK-1A and BI2536 in all oesophageal cell types (Figure

3-19, Figure 3-20). Comparing initial and final cell numbers after 96 hours of treatment, both PLK-1A and BI 2536 appear to induce a reduction of proliferation rather than true cytotoxicity.

To confirm a cytotoxic phenotype could be detected using the AlamarBlue method, cells were treated with the DNA-crosslinking agent Cisplatin. This drug forms the basis for the current neoadjuvant and palliative chemotherapy regimens used in oesophageal cancer and is cytotoxic to proliferating cells due to the formation of DNA-adducts and the induction of apoptosis[544]. Unfortunately, the therapeutic efficacy of this drug is limited by a number of side-effects including nephrotoxicity, neurotoxicity and ototoxicity[545]. All oesophageal cells tested exhibited dose-dependent cytotoxicity with Cisplatin treatment, providing evidence this method may allow the discrimination of cytotoxic and cytostatic drug effects (Figure 3-21).

When the relative responses of the cell lines were considered, cancer cell lines were more sensitive to PLK-1 inhibitors than HEEpiC cells (Table 3-6). The Barrett's cell line CP-A and the cancer cell lines demonstrate an equivalent sensitivity to BI2536 whereas the CP-A cell line was marginally less sensitive to PLK-1A than the cancer cells, although the relative differences in sensitivities are small. In contrast, the cancer cell lines and non-transformed cells exhibited similar sensitivities to Cisplatin, confirming non-selective activity of this drug.

Table 3-6 Comparison of IC50 Values Across Oesophageal Cell Types

	HEEpiC	CP-A	OE33	OE19	KYSE-30
BI2536 IC_{50} (nM)	23 (16-30)	12 (9-14)	8 (7-9)	12 (11-14)	13 (10-15)
BI 2536 Relative IC_{50}	1.0	0.50 (0.32-0.78)	0.36 (0.24-0.46)	0.54 (0.36-0.72)	0.55 (0.36-0.74
PLK-1A IC_{50} (nM)	154 (113-194)	110 (89-131)	78 (63-93)	65 (49-80)	74 (61-86)
PLK-1A Relative IC_{50}	1.0	0.71 (0.48-0.94)	0.50 (0.33-0.67)	0.42 (0.27-0.57)	0.48 (0.32-0.63)
Cisplatin IC_{50} (µM)	7.2 (4.2-10.2)	7.0 (5.0-8.9)	9.7 (6.3-13.0)	13.0 (4.8-21.2)	~2*
Cisplatin Relative IC_{50}	1.0	0.97 (0.69-1.24)	1.34 (0.88-1.81)	1.81 (0.67-2.94)	~0.30*

*Values in parentheses are 95% Confidence Intervals. *Approximate value only as poor quality curve fit*

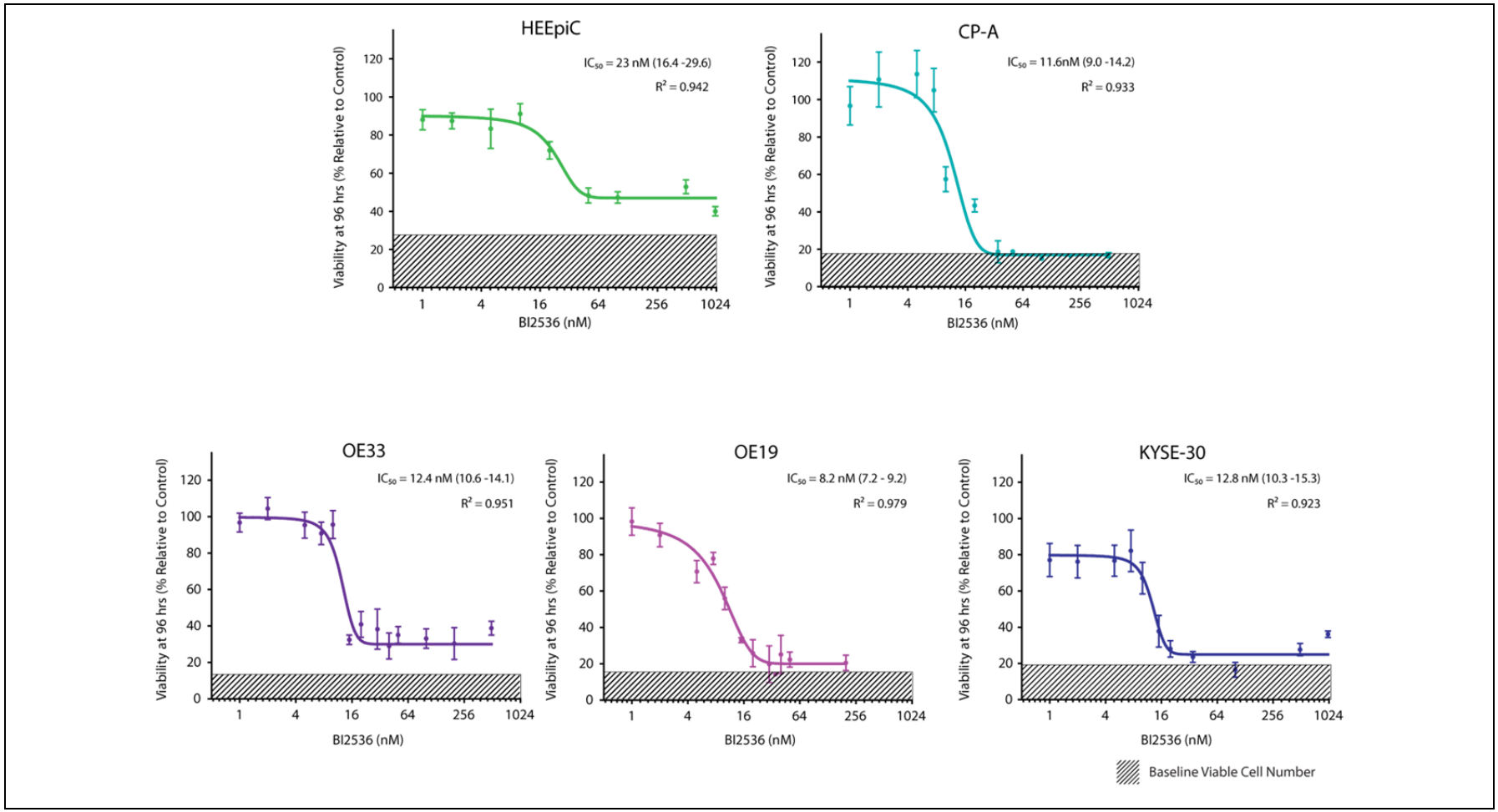

Figure 3-19 The Effect of BI2536 on Oesophageal Cell Viability. Cells were seeded in 96-well plates (HEEpiC – 10,000 cells per well, CP-A - 2,000 cells per well, OE19, OE33, KYSE-30 – 5,000 cells per well) and viability was measured after 24 hours (baseline viability). BI2536 dilutions or DMSO were added and viability was measured again at 96 hours after treatment. Points represent trimmed mean values (+/- SEM) from 4 biological replicates each consisting 8 technical replicates. Means were normalised to baseline viability and then DMSO treated controls. IC_{50} values, 95% confidence intervals (in parenthesis) and the R^2 for best fit curves are displayed.

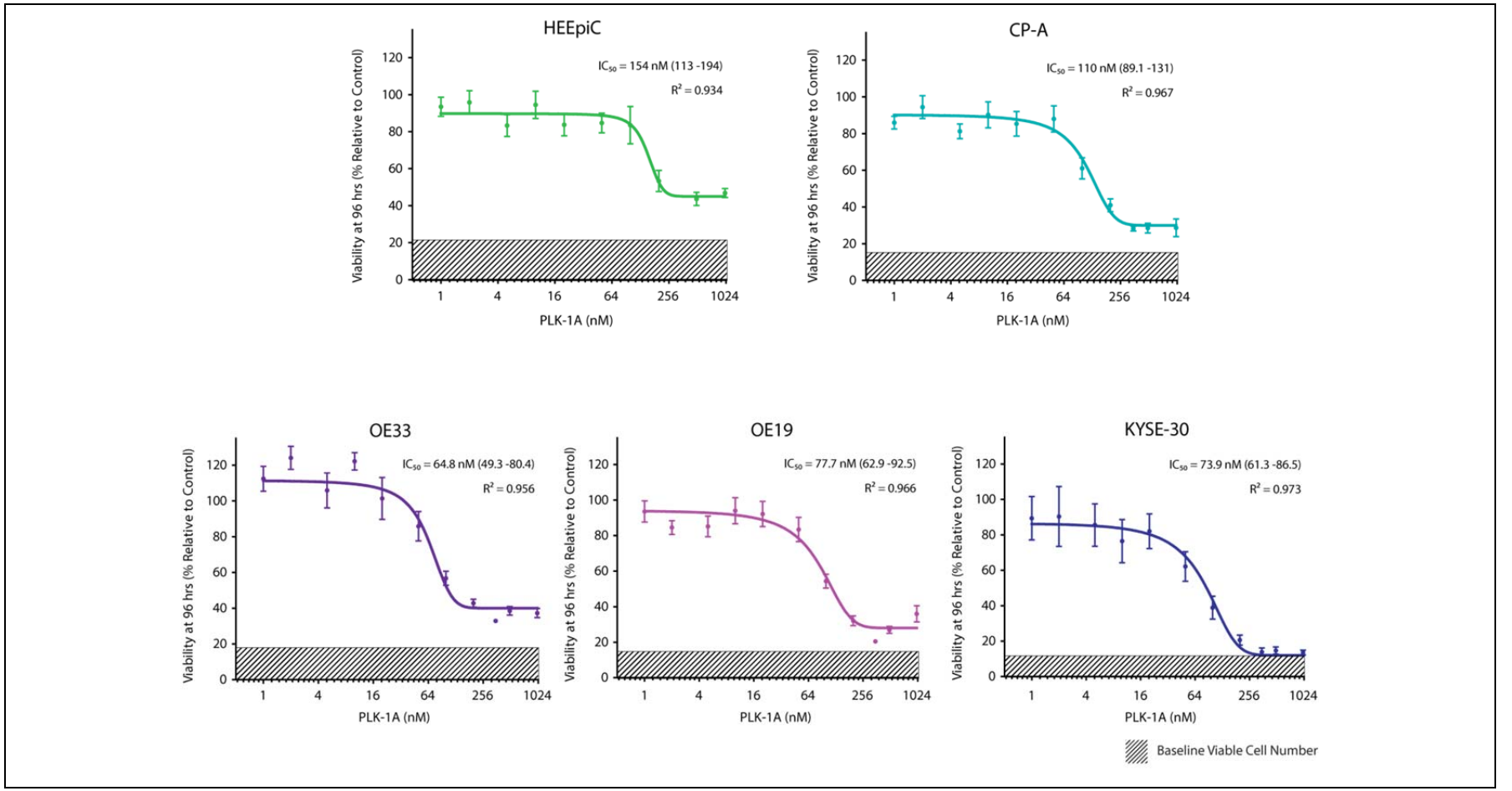

Figure 3-20 The Effect of PLK-1A on Oesophageal Cell Viability. Cells were seeded in 96-well plates (CP-A - 2,000 cells per well, HEEpiC – 10,000 cells per well, OE19, OE33, KYSE-30 – 5,000 cells per well) and viability was measured after 24 hours (baseline viable cell number). PLK-1A dilutions or DMSO were added and viability was measured again at 96 hours after treatment. Points represent trimmed mean values (+/- SEM) from 4 biological replicates each consisting 8 technical replicates. Means were normalised to baseline viability and then DMSO treated cells. IC_{50} values, 95% confidence intervals (in parenthesis) and the R^2 for best fit curves are displayed.

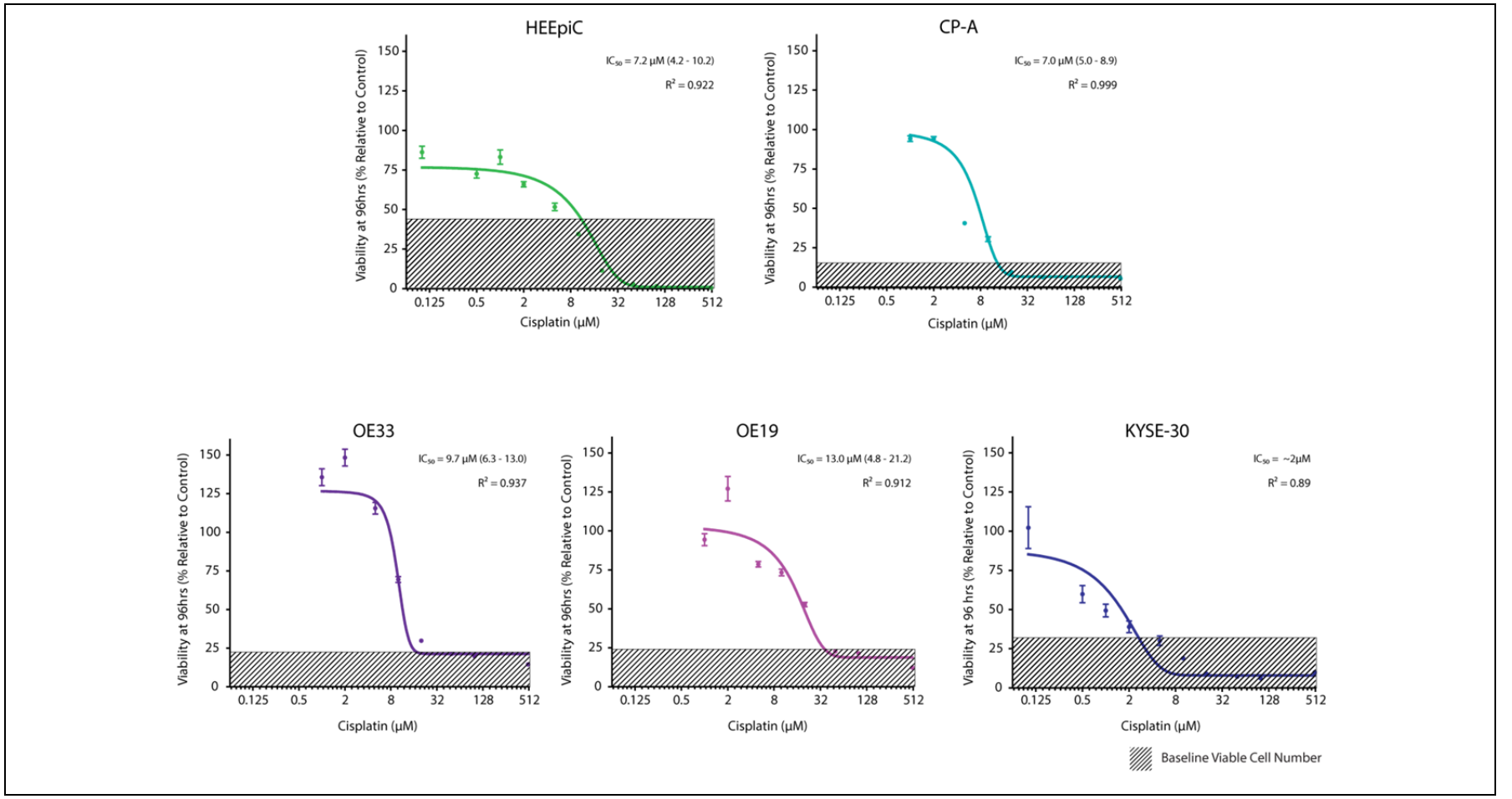

Figure 3-21 The Effect of Cisplatin on Oesophageal Cell Viability. Cells were seeded in 96-well plates (HEEpiC – 10,000 cells per well, CP-A - 2,000 cells per well, OE19, OE33, KYSE-30 – 5,000 cells per well) and viability measured after 24 hours (baseline viable cell number). Cisplatin dilutions or PBS were added and viability was measured again at 96 hours after treatment. Points represent trimmed mean values (+/- SEM) from 1 experiment consisting 8 technical replicates. Means were normalised to baseline viability and then DMSO treated controls. IC_{50} values, 95% confidence intervals (in parenthesis) and the R^2 for best fit curves are displayed.

These data suggest oesophageal cancer and Barrett's cells may be more sensitive to PLK-1 inhibition than HEEpiC cells. The therapeutic window, although narrow on the basis of the AlamarBlue data alone, suggests that tumour-specific effects may be achievable with careful dosing schedules. The dose-responses observed are in the constant presence of the inhibitor. This phenotype may not be observed, however, with the fluctuating tumour drug concentrations observed *in vivo*.

A colony formation assay was performed to investigate if the therapeutic window was altered after drug removal. Cells were incubated in PLK-1 inhibitor or vehicle for 72 hours, trypsinised and seeded in fresh media at several densities (Figure 3-22). This revealed complete inhibition of colony formation after treatment with 100 nM BI2536 in OE33 and CP-A cells and in contrast, no appreciable reduction in colony formation at 10 nM. This pattern of sensitivity closely correlates with the AlamarBlue dose-response curves (Figure 3-19). The OE19 cell line also showed a significant reduction in colony formation at the higher BI 2536 dose and only a small reduction in colony number at 10 nM (Figure 3-22). In contrast the OE33 cells appeared more sensitive to PLK-1A treatment than CP-A cells with OE19 cells showing an intermediate response.

The complete inhibition of colony formation is surprising as the AlamarBlue dose-response curves suggested both inhibitors significantly reduce but do not abolish proliferation. In the colony formation assay, all of the plates were seeded with viable cells as assessed by trypan blue exclusion. Even in the absence of proliferation, some staining would be expected. The combination of trypsinisation and drug removal appears to induce cell death.

To investigate if drug removal alone reduces viable cell number, the AlamarBlue method was adapted for a death commitment assay (DCA). In this assay, AlamarBlue measurements were taken before drug addition, after 24 hours of drug exposure and again 72 hours after either continued drug exposure or after drug withdrawal. If fewer viable cells were present at the end compared to the interim point this suggested cell death was occurring[546].

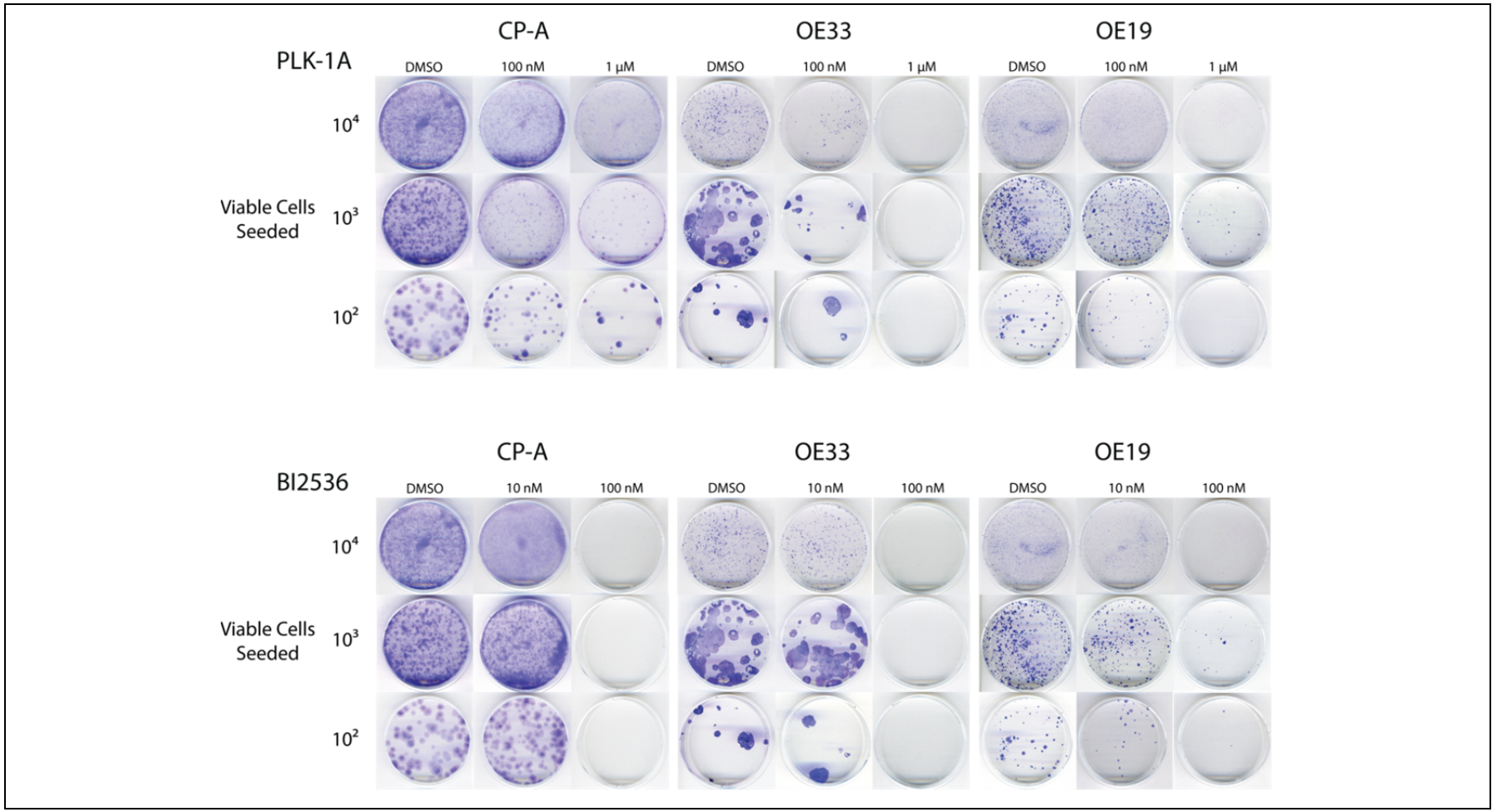

Figure 3-22 Colony Formation Assay After PLK-1 Inhibitor Treatment. Cells were treated with a PLK-1 inhibitor or vehicle at the doses shown and then trypsinised and equal numbers of viable cell seeded in fresh media. Plates at each seeding density were fixed and stained for colonies when vehicle treated cells approached confluence. Seeded viable cell number; A = 1x10^4 (CP-A, OE33), 5x10^4 (OE19); B = 1/10 dilution of A; C = 1/100 dilution of A. Images representative of 2 independent experiments.

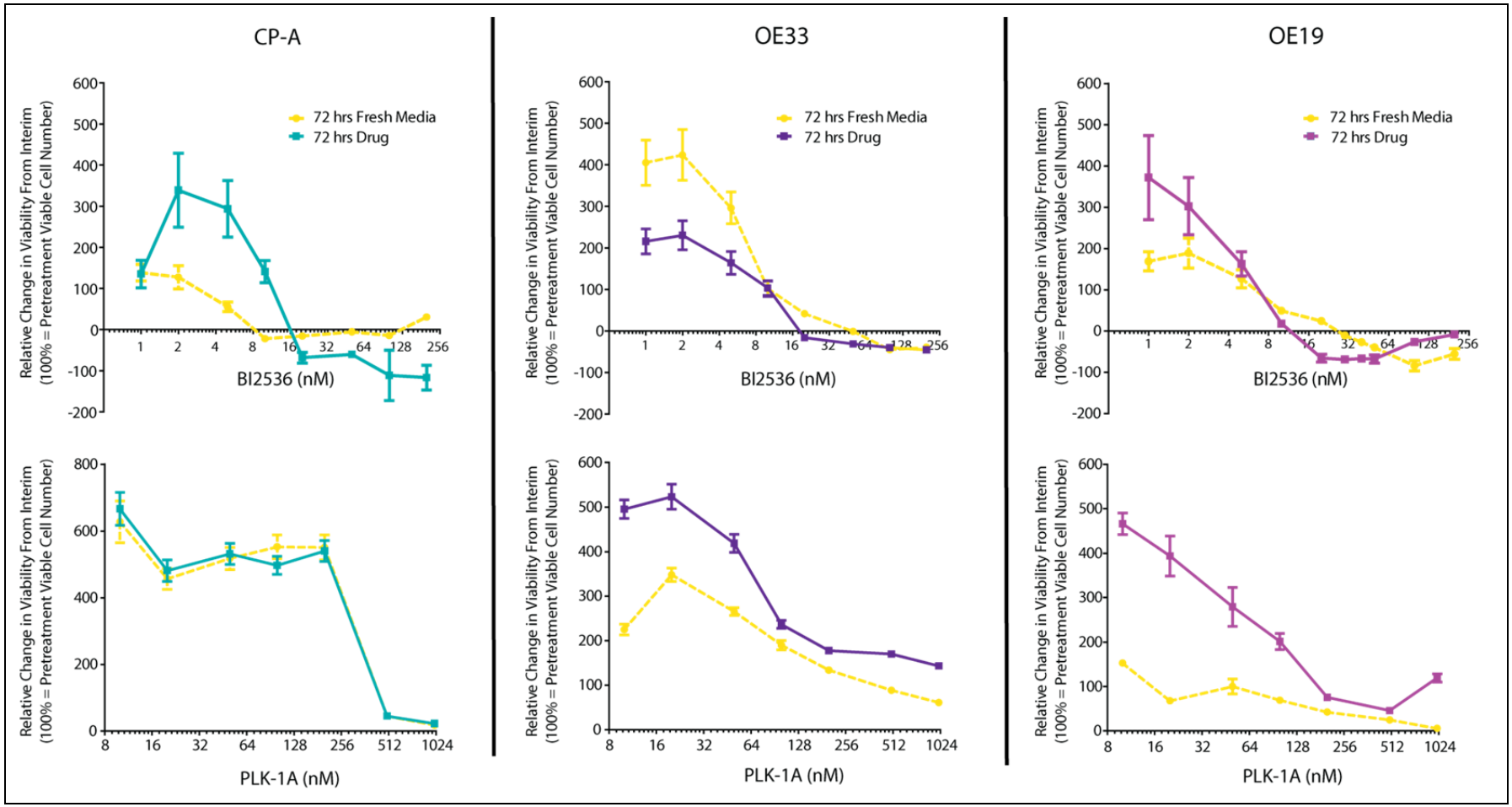

Figure 3-23 Death Commitment Assay After 24 hours PLK-1 Inhibitor Treatment. Cells were treated with BI2536 or PLK-1A for 24 hours and then either treated with fresh media or drug at the same concentration for a further 72 hours. Points represent trimmed mean change in viability from the interim point (after 24 hours drug treatment) +/- SEM from 8 technical replicates from 2 independent experiments. Values have been normalised so 100% on the Y-axis represents the pre-treatment viable cell number.

After treatment with PLK-1A for 24 hours, further drug treatment did not change cell proliferation compared to media alone in CP-A cells (Figure 3-23). In contrast in OE33 and OE19 cells, pulsed PLK-1A exposure appeared to enhance the anti-proliferative effect compared to persisting treatment.

The results of treatment with BI2536 in the DCA show some similarity with those of PLK-1A. At higher doses with both treatments, there is little additional cell growth in any of the oesophageal cells after 24 hours drug exposure regardless of subsequent treatments. The results from treatment with lower doses of BI2536 appear to show proliferation in all the cell types with an enhanced anti-proliferative effect after pulsed exposure in the CP-A and OE19 cells and reduced effect in the OE33 cells.

Real-time imaging of PLK-1 inhibitor uptake and removal in cells is not currently possible. To explore if the wash-off protocol was adequate to allow drug removal, a compound that exhibited a reversible growth phenotype was required. The MDM2 antagonist Nutlin-3A has been reported to induce a reversible growth arrest in epithelial cancer cell lines with wt p53 but not those with mutant p53[547,548]. This is thought to be secondary to selective p53-induction due to the blockade of the MDM2-p53 interaction which usually leads to p53 degradation. The effect of Nutlin-3A on viability was examined in oesophageal cells.

This revealed dose-dependent cytotoxicity in all three cell types with a 6 fold and 2.5 fold greater sensitivity in the CP-A (wt p53) cells than OE33 and OE19 cells (both mutant p53) respectively (Figure 3-24). The death commitment assay revealed proliferation to a similar extent in all cells when washed and grown in fresh medium, irrespective of the prior dose of Nutlin-3A. This suggests the washing protocol in the DCA is sufficient to remove a drug and allow growth recovery.

Given the proliferation of cells after Nutlin-3A removal, the enhanced proliferation of CP-A and OE19 cells treated continuously with low doses Nutlin-3A compared to those treated with pulsed Nutlin-3A is surprising. This suggests either low dose Nutlin-3A promotes growth or, as previously, there may be a systematic bias towards reduction in observed viable cell number with pulsed drug exposure in the DCA.

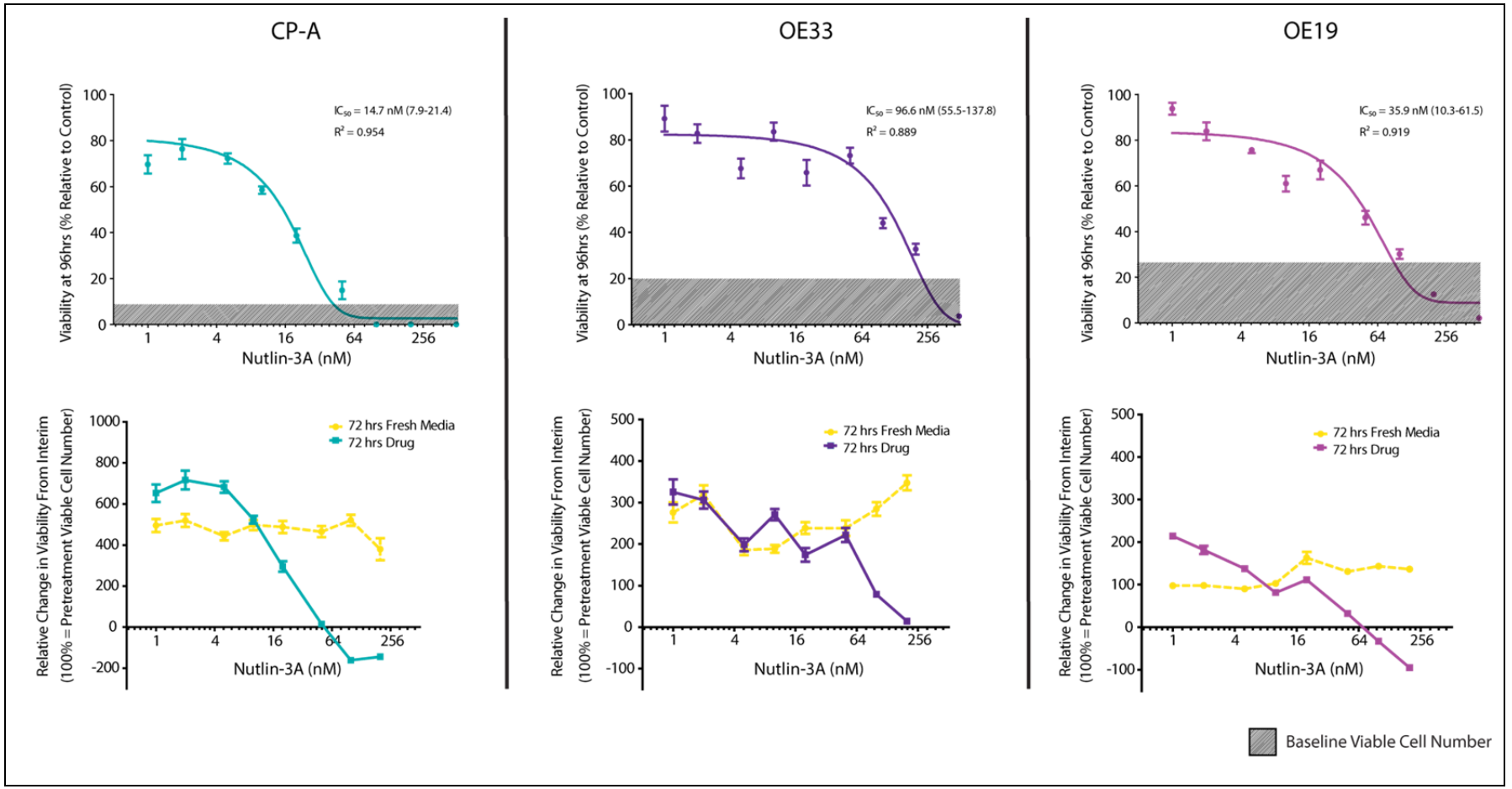

Figure 3-24 Effect of Nutlin-3A on Oesophageal Cell Viability. Lower figures – Cells were treated for 24 hours and then a Death Commitment Assay (DCA) was performed in an identical fashion to Figure 3-23. Points represent trimmed mean change in viability from the interim point, 24 hours after drug treatment, +/- SEM from 8 technical replicates in one experiment. Values have been normalised so 100% on the Y-axis represents the pre-treatment viable cell number. Upper figures – the baseline and final viability data from cells treated for 96 hours with Nutlin-3A or DMSO as part of the DCA are presented as dose-response curves (calculations identical to Figure 3-19).

To establish if changing culture media creates an artefact that can explain the apparent cell death commitment after pulsed drug administration, real-time monitoring of cell number was employed using an orthogonal cell proliferation assay (xCELLigence). Initial studies revealed greater sensitivity for changes in cell number using cell lines that grew with an adherent epithelial morphology. Therefore the results of pulsed PLK-1 inhibitor exposure on cell proliferation were determined with the CP-A and KYSE-30 cell lines because of their adherent growth pattern.

The process of removal from the incubator, drug addition or wash-off and plate replacement on the RTCA DP instrument led to significant changes in impedance readings that stabilized after 1-2 hours further incubation. These could be secondary to cells being detached from the culture surface, fluctuations in the plate temperature and/or contact area between the plate and the impedance monitor. To prevent these changes influencing the subsequent results, readings were normalized to a point between one and two hours after drug addition or wash-off when cell index measurements had stabilized. Therefore change in cell number caused by detachment during washing would be accounted for and only subsequent changes in cell number due to proliferation or cell death would affect the normalized cell index.

A caveat with the xCELLigence technique is that changes in cell morphology, including those that are expected with drugs that block cells in mitosis or at the G2/M checkpoint, change impedance. Therefore, although unlikely, the results may represent drug-induced cell morphology changes rather than effects on proliferation.

To confirm xCELLigence cell index measurements represent viable cell number changes, a comparison was made with another viability assay. The E-plate wells are lined with electrodes and therefore fluorescence or colorimetric measurements, including the AlamarBlue method, cannot be used. Cells can be lysed in the E-plates with CellTiter-Glo reagent and the lysate transferred to white opaque 96 well plates without significant change in relative viability measurements (Figure 3-12A). The final cell index after a 96 hour treatment with BI2536 was compared to the CellTiter-Glo measurement obtained from the same wells (Figure 3-25).

Comparison between relative viable cell number estimates from the xCELLigence cell index and CellTiter-Glo revealed an excellent correlation (CP-A, Pearson R^2 = 0.9602, $p<0.0001$ and KYSE-30, Pearson R^2 = 0.9918, $p<0.0001$). This confirms xCELLigence cell index is a reliable arbiter of viable cell number and could be used to monitor cell proliferation in real-time.

To allow a reasonable period of cell recovery after drug wash-off without DMSO treated wells becoming confluent, a small starting number of cells was used for both cell lines. For the CP-A cell line, 1,000 cells was the minimum number that was compatible with subsequent proliferation. The KYSE-30 cell line showed a slightly more rapid growth rate than CP-A cells (Figure 3-26A, Figure 3-27A) and approximately 250 cells were used per well.

PLK-1A produced a dose-dependent reduction in proliferation, in both cell lines in keeping with previous results. The small number of KYSE-30 cells meant changes in cell index were small with greater variability in the dose response determination (Figure 3-27B, D). In both cell types, however, there were clear changes in the growth pattern of cells in the presence of PLK-1A and after PLK-1A wash-off. CP-A cells proliferated in 200 nM PLK-1A but cell death was observed after drug wash-off (Figure 3-26B, C). The effect was even more marked in the KYSE-30 cell line with cells showing proliferation in 200 nM and 500 nM PLK-1A and cell death after drug wash-off (Figure 3-27 B, C).

When the growth rate per hour was normalised to DMSO treated cells, there was a statistically non-significant decrease in the IC_{50} for CP-A cells from 190 nM to 145 nM after drug wash-off. In contrast the IC_{50} showed a significant decrease from 430 nM to 70 nM (Figure 3-27 D, E; $p = 0.0188$, two-tailed t-test) in KYSE-30 cells. This suggests an enhanced therapeutic effect for this drug *in vitro* with pulsatile rather than continuous exposure.

The response of CP-A cells to BI2536 was also assessed using the xCELLigence method in the same fashion as PLK-1A (Figure 3-28). This revealed a similar pattern to previous results obtained using the AlamarBlue method with dose-

dependent inhibition of proliferation. The IC_{50} for reduction of proliferation when cells were incubated with the drug was unchanged when the drug was washed off, although cell death was observed at higher drug concentrations after drug removal. Therefore it appears that pulsed administration of PLK-1 inhibitors does not significantly enhance or reduce the anti-proliferative effect in CP-A cells but may augment the anti-proliferative effect and induce cell death in oesophageal cancer cells.

To evaluate if a shorter pulse of treatment with PLK-1 inhibitors could enhance the therapeutic window between CP-A and adenocarcinoma cell lines, a colony outgrowth assay was performed (Figure 3-29).

This revealed dose-dependent inhibition of colony growth in all cell lines. There was a small reduction in colony growth in OE33 cells after 6.3 nM BI 2536 treatment and no appreciable effect in CP-A or OE19 cells. With 25 nM BI 2536 treatment, there was significant inhibition of colony growth across all cell lines. This is in close agreement with the AlamarBlue dose-response curves (Figure 3-19) and confirms a narrow dose range between no effect and maximal effect for this drug. This also suggests little change in the sensitivity of oesophageal cells to this drug with pulsed rather than constant administration in agreement with the real-time proliferation data (Figure 3-28).

Similar to previous findings with the colony formation assay, OE33 cells showed enhanced sensitivity to the PLK-1A compound compared to CP-A and OE19 cells. Reduced colony outgrowth was observed in OE33 cells after pulsed administration of 16 nM PLK-1A with no effect in CP-A and OE19 cells and there was almost complete eradication of OE33 colony growth after 63 nM PLK-1A treatment with only a small reduction in colony growth in CP-A and OE19 cells. This is in contrast to findings with the AlamarBlue assay where persistent proliferation was observed with 50 -100 nM PLK-1A treatment in all three cell lines (Figure 3-20).

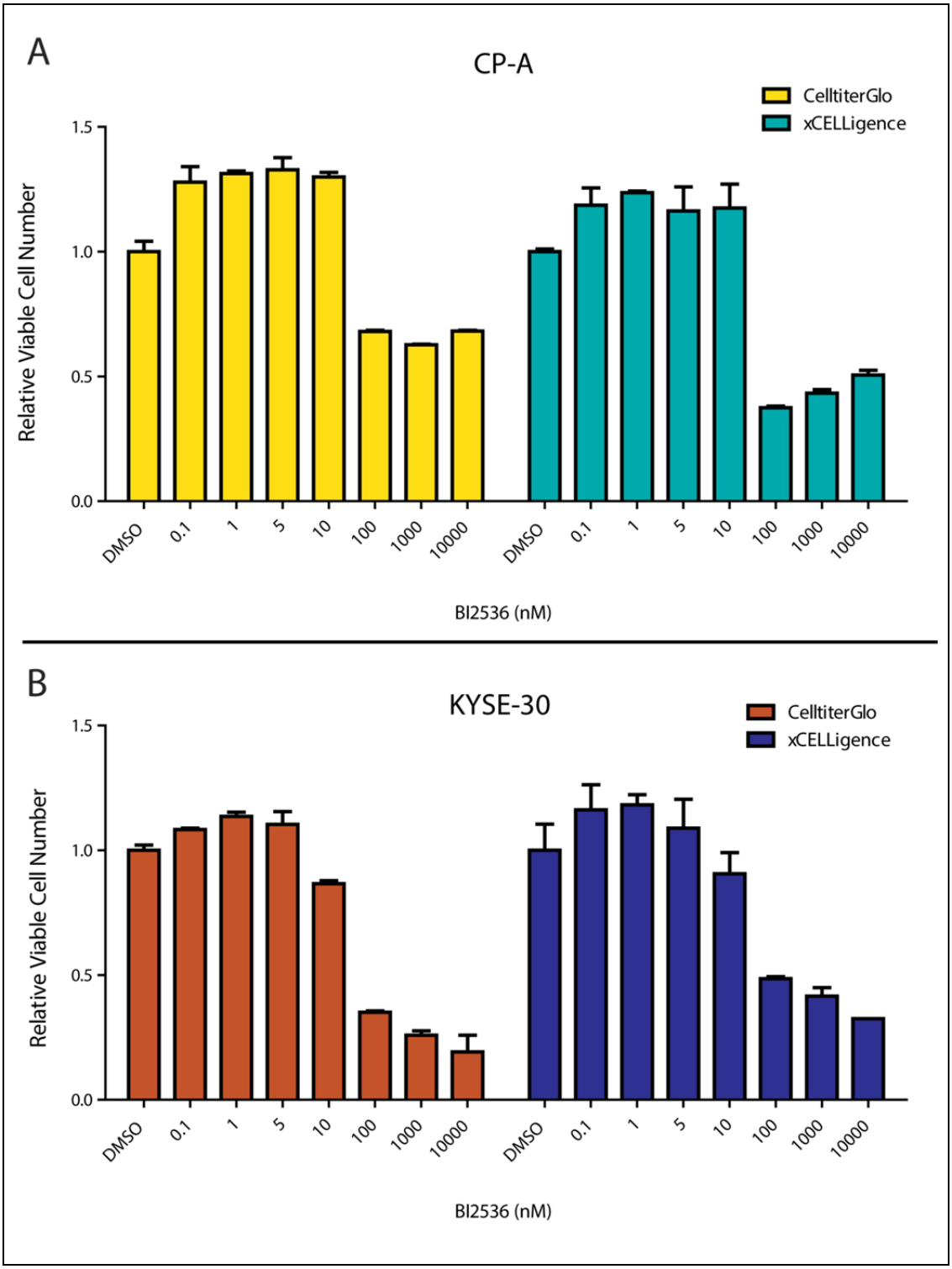

Figure 3-25 Comparison of Viable Cell Number Estimation by xCELLigence and CellTiter-Glo. (A) After 24 hours incubation on E-plates, (A) CP-A or (B) KYSE-30 cells were treated with BI2536 or DMSO and final cell index calculated, 96 hours after drug addition. CellTiter-Glo measurements were obtained from the same wells. Bars represent mean +/- SEM of technical duplicates normalised to DMSO treated cells.

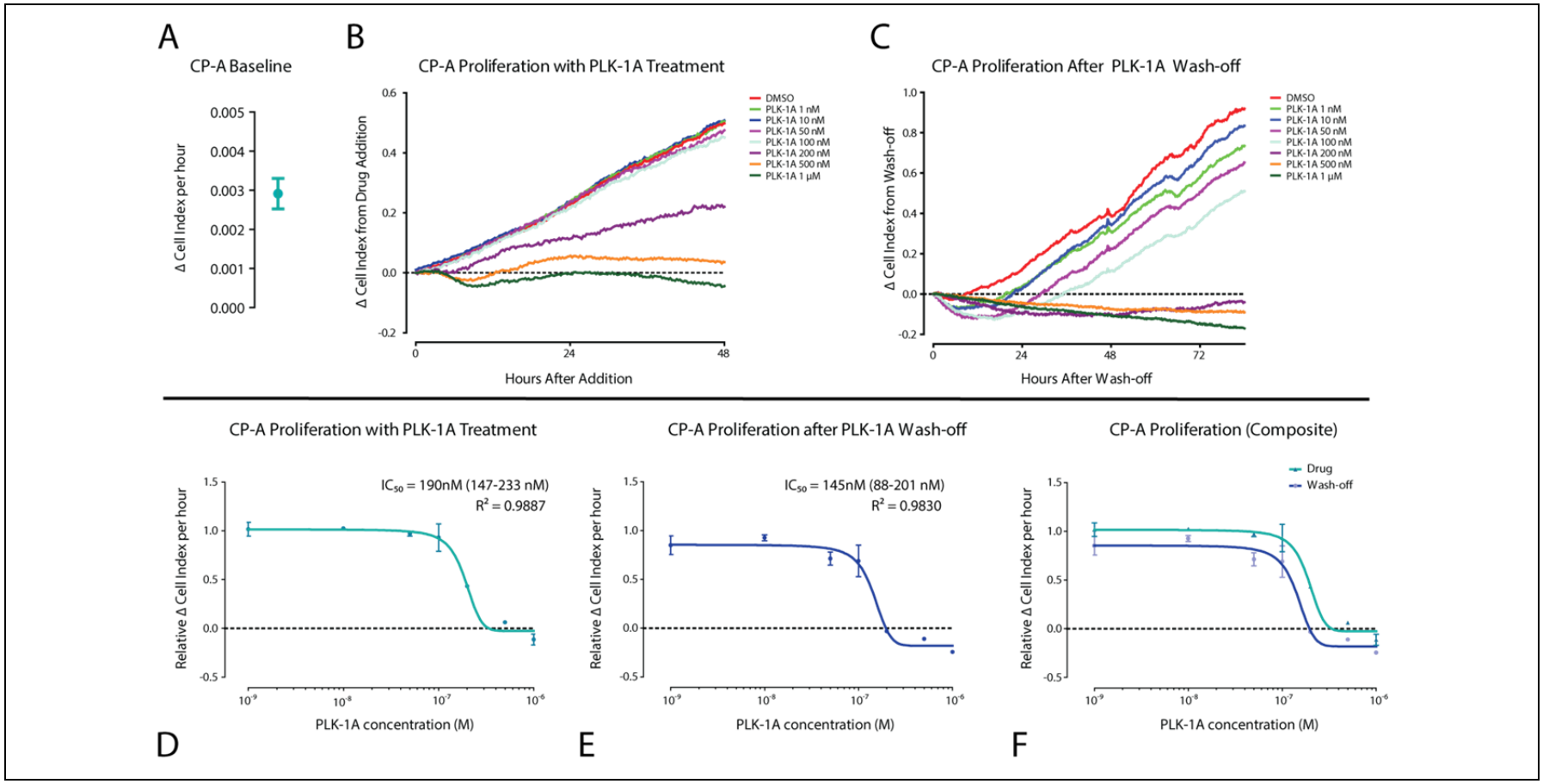

Figure 3-26 Real-time Monitoring of CP-A Proliferation after PLK-1A Treatment. (A) Cells (1,000 per well) were seeded and allowed to proliferate for 24 hours. The mean change in cell index per hour across the wells +/- SEM is displayed. (B) PLK-1A was then added and proliferation monitored continuously as the change in cell index over 48 hours. Proliferation continued to be monitored for 90 hours after the drug was washed off (C). (B, C) Lines represent mean of technical duplicates; values less than zero represent a decrease in cell number. (D, E) The change in cell index per hour in the presence of drug (D), or after drug wash-off (E) normalised to DMSO treated cells. Points represent mean +/- SEM from technical duplicates. IC_{50} values, 95% CI (in parenthesis) and the R^2 for best fit curves are displayed. (F) Composite of (D) and (E).

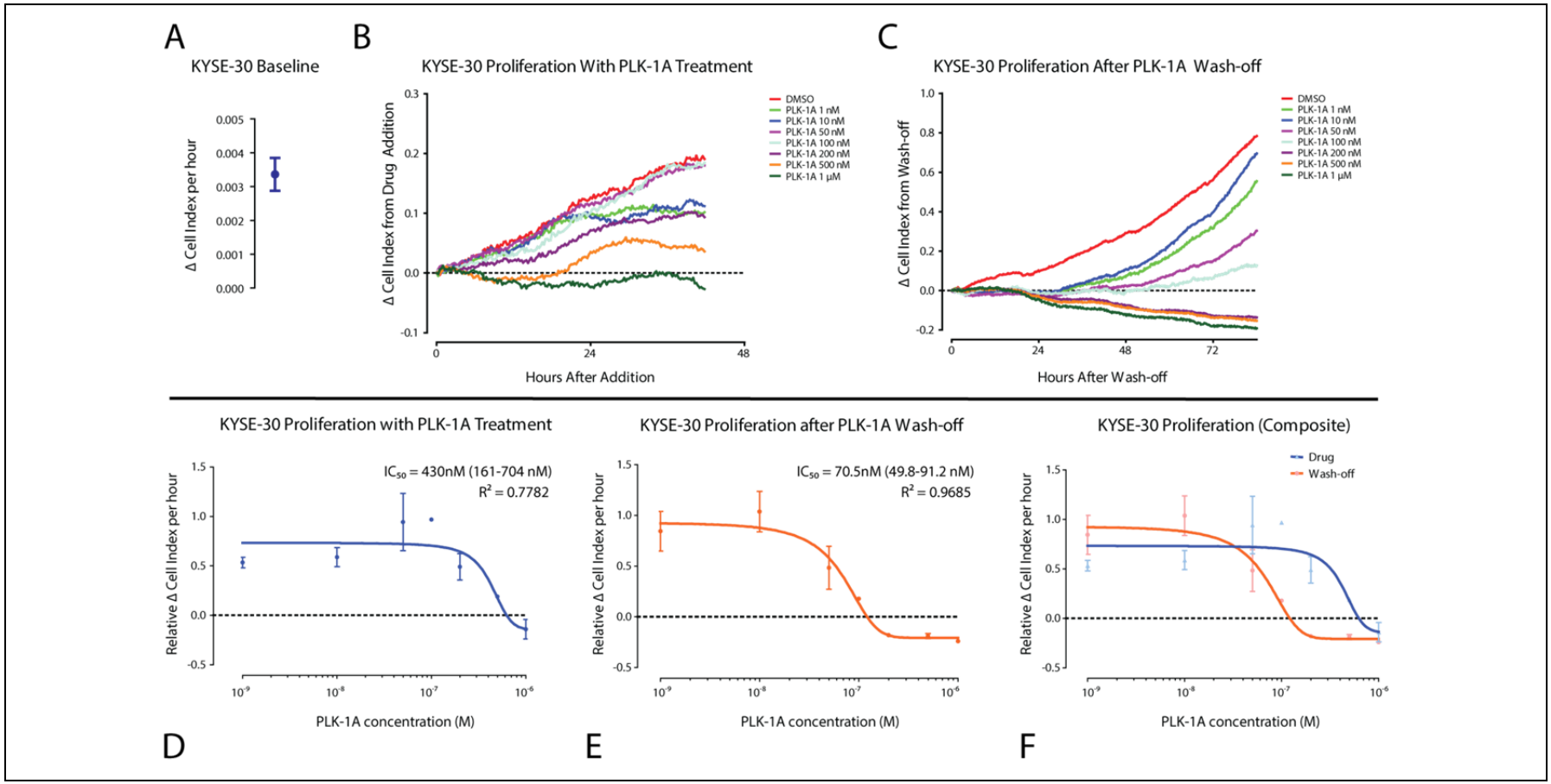

Figure 3-27 Real-time Monitoring of KYSE-30 Proliferation after PLK-1A Treatment. (A) Cells (250 per well) were seeded and allowed to proliferate for 24 hours. The mean change in cell index per hour across wells +/- SEM is displayed. (B) PLK-1A was then added and proliferation monitored continuously as the change in cell index over 42 hours. Proliferation continued to be monitored for 90 hours after the drug was washed off (C). (B, C) Lines represent mean of technical duplicates; values less than zero represent a decrease in cell number. (D, E) The change in cell index per hour in the presence of drug (D), or after drug wash-off (E) normalised to DMSO treated cells. Points represent mean +/- SEM from technical duplicates. IC_{50} values, 95% CI (in parenthesis) and the R^2 for best fit curves are displayed. (F) Composite of (D) and (E).

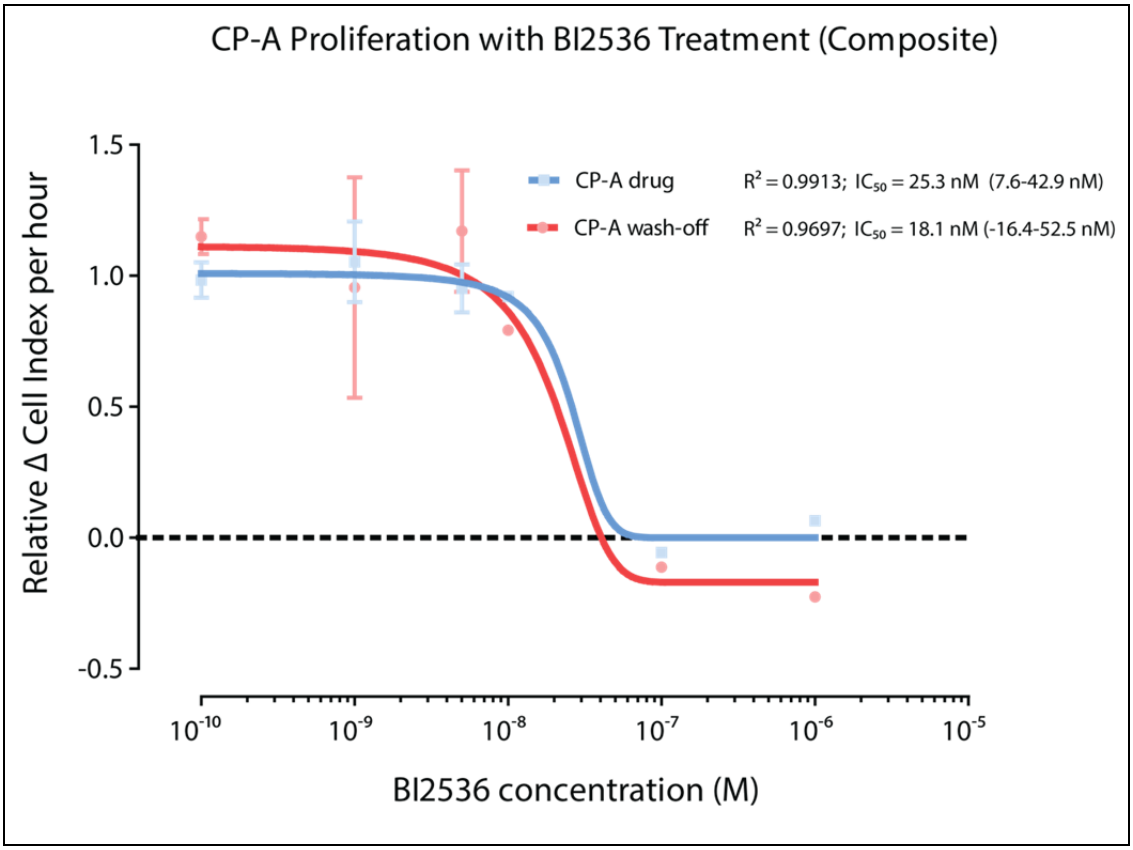

Figure 3-28 Real-time Monitoring of CP-A Proliferation after BI2536 Treatment. Cells (1,000 per well) were seeded and allowed to proliferate for 24 hours. BI2536 or DMSO was added and proliferation monitored continuously as the change in cell index over 48 hours. The drug was then washed off and proliferation monitored for a further 96 hours. The change in cell index per hour in the presence of drug (blue points, drug), or after drug wash-off (red points, wash-off) normalised to DMSO treated cells is presented. Points represent mean +/- SEM from technical duplicates. IC_{50} values, 95% CI (in parenthesis) and the R^2 for best fit curves are displayed.

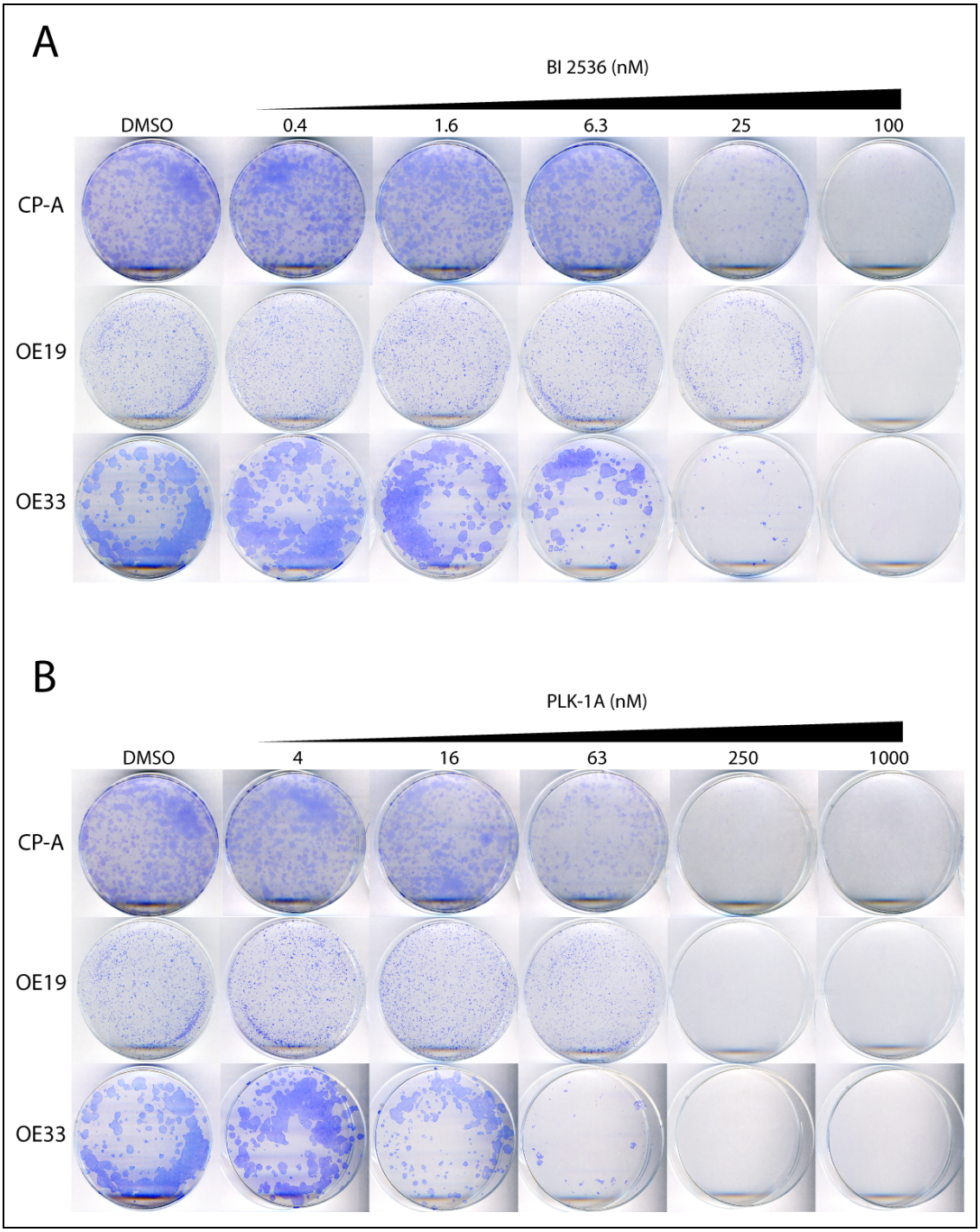

Figure 3-29 Colony Outgrowth Assay. Cells were plated in complete medium (CP-A, OE33 – 2,500 cells per plate; OE19 – 10,000 cells per plate), allowed to adhere and then treated for 16 hours with drug or DMSO followed by incubation in complete medium. Plates were fixed and stained for colonies when vehicle treated cells approached confluence (12 days after drug wash-off).

3.4 Discussion

Both STR genotypes and p53 functional status were used to verify the oesophageal cell lines used in this study. Using these *bona fide* cell lines, PLK-1 was demonstrated to be over-expressed in the transformed compared to non-transformed cells. Higher expression of PLK-1 has been previously demonstrated between OSCC cell lines and HEEpiC cells using a different anti-PLK-1 antibody (clone 36-298)[270]. The specificity of the antibody used in this study was confirmed by siRNA and western blotting.

Transfection of non-targeting scrambled siRNA and siRNA to GAPDH produced a reduction in PLK-1 expression. This may either be due to a direct interaction with PLK-1 transcripts, although this is far less likely from both scrambled and siGAPDH sequences, or a consequence of secondary effects of siRNA on cell cycle progression or other cellular processes linked to mitosis. A change in cell cycle stage could have significant effects on PLK-1 expression as transcript levels have been shown to peak in OAC cells just prior G2/M phase[276].

The expression of PLK-1 in the cell line panel could be explained by differential rates of proliferation. The oesophageal cells tested exhibit *in vitro* doubling times of between 18 and 32 hours. This is in stark contrast to *in vivo* human tumours which, in many cancer types, exhibit doubling times of several weeks to months with only 1-2% of cells dividing at any point[549]. When mRNA expression has been measured, however, PLK-1 has been found to be upregulated in OAC and OSCC cell lines and tumours compared to normal squamous cells and tissue[271,275,276]. Overexpression of PLK-1 in oesophageal tumours has also been associated with poorer prognosis and increased risk of lymphatic metastasis[268,275].

PLK-1 was shown to be overexpressed in OAC tissue compared to Barrett's and normal squamous epithelium by IHC. This provides some evidence that the cell line panel may recapitulate some of the features of the *in vivo* tissues. Loss of p53 function may be responsible for the upregulation in PLK-1 as this has been demonstrated previously *in vitro* [220,340]. Supporting this, the oesophageal cell lines with the highest PLK-1 expression possess p53 mutations. These cancer cell lines

also possess other genetic changes relative to the CP-A and HEEpiC cells that together create the transformed phenotype. No oesophageal adenocarcinoma cells lines are available with wild-type p53 to assess the direct impact of loss of p53 on PLK-1 expression in this cell type.

Assessing PLK-1 expression across a large panel of oesophageal adenocarcinomas with both wild-type and mutant p53 may allow the impact of this genetic change on PLK-1 expression to be determined. There are a myriad of genetic changes in these tumours but, with a large enough cohort, it would be expected that those features not correlated with TP53 mutation will be equally represented in tumours with wild-type and those with mutant alleles.

Using the OCCAMS TMA consisting of cores from 356 OACs, TP53 genotype and PLK-1 expression were assessed. Using a sub-group of 58% of the tumours, TP53 mutation was unanimously predicted in 90% of the tumours. This high figure in part reflects the greater ease in identifying mutant p53 tumours compared to wild-type p53 tumours using p53 IHC alone. A p53 mutation frequency of 72% was, however, identified in a recent exome-sequencing study of 149 oesophageal adenocarcinomas from American patients[550]. A further amplicon re-sequencing study in 99 patients supports these findings with a mutation frequency of 71% (Prof. Rebecca Fitzgerald, personal communication). In support of the IHC method, the sequencing data from a small selection of the TMA cases largely confirmed the IHC predictions. Definitive TP53 sequencing of the entire cohort would have clarified the accuracy of the predictions. This would require DNA extraction from the FFPE tissue, further reducing the availability of this highly limited resource and re-sequencing the entire TP53 gene in over 300 tumours would have been prohibitively expensive.

PLK-1 expression was found to be significantly elevated in oesophageal adenocarcinomas with predicted mutant TP53 compared to those tumours with predicted wild-type p53. This was in agreement with the previous findings comparing non-transformed oesophageal cells (wild-type p53) and oesophageal cancer cell lines (mutant p53) and benign and malignant oesophageal tissues (Figure 3-4, Figure 3-9). These findings were reproduced in ovarian cancers with known

TP53 genotype. This suggests the previous findings in OAC were not simply an artefact of the method used to predict TP53 genotype. Together these data suggest mutation of p53 is associated with upregulation of PLK-1 *in vivo*.

To provide definitive proof for the association between TP53 function and PLK-1 expression in oesophageal cancer, a panel of isogenic cell lines would be required and these have not yet been developed. Another possibility would be to replace p53 function in an OAC cell line with a mutant allele and evaluate PLK-1 expression. Unfortunately many mutant p53 isoforms exhibit dominant negative effects by binding to and inactivating wild-type proteins[202,551]. Exogenous expression of wild-type protein may not, therefore, rescue p53 function. A further complicating factor would be the requirement for careful calibration of wild-type p53 expression as supra-normal p53 activity is associated with diverse physiological changes *in vivo* including premature aging and resistance to cancer development[552,553].

To examine if p53 mutation and subsequent higher PLK-1 expression conferred greater sensitivity to PLK-1 inhibition, both siRNA to PLK-1 and PLK-1 inhibitors were tested in transformed and non-transformed oesophageal cells with measurement of viability. The AlamarBlue assay was used because this exhibited a linear response with increasing cell number and initial viability measurements did not influence final viable cell number determination.

Viability experiments were optimised for 96-well plates. After siRNA transfection to PLK-1, insufficient cells were present at the end of the experiment to allow subsequent RNA expression profiling or protein expression analysis by western blotting. AlamarBlue assays allow downstream assays to be performed including the In-Cell western method[554]. When this method was applied with CP-A and OE19 cells, however, a high coefficient of variation in the determination of PLK-1 expression was observed. Subsequently, parallel rather than serial assays of viability, gene and protein expression were performed.

When scrambled siRNA sequences (siGENOME, Non-Targeting siRNA#3, Dharmacon) were transfected into oesophageal cells, PLK-1 mRNA was reduced

(Figure 3-17). This is likely to be a sequence specific phenomenon as it was observed to a similar extent in all 3 cell types and not observed in vector treated cells or when a different sequence was used (OE33, HEEpiC cells; Accell, non-targeting sequence #1, Dharmacon). To mitigate the off-target effects of siRNA, individual sequences could be transfected as pools of 4 sequences were used. This may eliminate the non-specific interacting sequence. Failing that, the use of lentiviral vectors to generate stable, inducible shRNA expressing cell-lines may prevent non-specific effects although this would be at the expense of significantly increased generation time.

Despite the effects of the non-targeting control on mRNA expression, viability was unchanged, with the exception of HEEpiC cells, between vector and non-targeting control treated cells. This suggests, as expected, it is only the reduction in PLK-1 protein expression that is associated with a reduction in viability.

It was difficult to directly correlate absolute protein levels and viability but the cell line with the lowest reduction in PLK-1 protein expression, OE33, also exhibited the lowest decrease in viability. No clear discrimination between transformed and non-transformed cells could be determined from the siRNA experiment.

Two PLK-1 inhibitors were assessed, BI2536 and a novel, shorter half-life, compound PLK-1A. Both compounds produced the expected G2/M arrest after 24 hours and a reduction in proliferation was observed after 96 hours treatment in a both non-transformed and transformed cells. The degree of selectivity of a drug for cancer cells over normal and haematopoietic tissues can be considered as the therapeutic index. Cell lines derived from the same tissue, with a presumed highly similar genetic background, except for key transformative changes, have been used in this investigation in an attempt to reveal cancer cell selectivity. It is hypothesised that drugs affecting only the oesophageal cancer cell lines would exhibit a high therapeutic index *in vivo.*

OAC cell lines were twice as sensitive to the PLK-1 inhibitors as primary squamous cells although in the presence of the drug, cell numbers continued to increase even at

the highest doses tested. In contrast, cisplatin produced the expected cytotoxic response with minimal discrimination between cancer and non-transformed cells. The response to PLK-1 inhibitors could be due to a predominantly cytostatic rather than cytotoxic action of the drugs. A combination of proliferation and cell death could, however, produce the same results and real-time monitoring would be required to confirm this.

The colony formation and DCA were used to assess the effect of drug removal on subsequent viability. After treatment with PLK-1A, OAC cells showed a greater reduction in proliferation in the DCA after drug wash-off compared to CP-A cells. There was some similarity with the BI2536 treatment although significant variability was noted and further biological replicates would be required before the significance of these results can be determined.

A possible explanation for these data is that the wash-off has some effect in removing the drug and that this triggers a differential response in the adenocarcinoma cells. This is supported by the enhanced sensitivity of OE33 cells to PLK-1A observed in the colony formation assay (Figure 3-22). The persisting increase in viable cell number after pulsed and continuous PLK-1A treatment agrees with previous observations that this treatment has a net cytostatic, rather than cytotoxic, effect in the time scale of this assay (Figure 3-20). The lack of viable colonies after 1 μM PLK-1A treatment in OE33 cells could be due to delayed cell death (Figure 3-22).

In the DCA, cells at the interim viability assay stage have been pre-treated for 24 hours in conditions that have been previously shown to significantly induce G2/M arrest (Figure 3-18). As cells approach mitosis, there is a change in morphology that would be expected to reduce cell adhesion and therefore facilitate detachment during washing[555]. A further possibility to explain the reduced apparent proliferation after pulsed drug exposure would be that cells have been washed off after the interim viability assay and a reduced final viable cell number is just a consequence of the washing steps. The same number of washing steps was, however, employed for both pulsed and continuous drug exposed cells and it is reasonable to expect the effect to

be constant across groups. A systemic bias towards cell detachment and the death of detached cells with pulsed exposure could, in theory, explain the findings of an apparent enhanced anti-proliferative effect after drug removal.

The single wash after the interim AlamarBlue measurement may also be insufficient to wash-off the PLK-1 inhibitors and the results observed may not reflect the *in vivo* phenotype after pulsed administration. The wash protocol appeared, however, to be adequate to reverse the phenotype of Nutlin-3A.

A surprising result from the DCA was the cytotoxic effect of Nutlin-3A on cells with mutant p53. It is expected that Nutlin-3A would not change the phenotype of OE33 cells as mutant p53 is stabilised and non-functional (B). Similarly, OE19 cells possess a frameshift p53 mutation which produces a truncated protein which is predicted to lack the tetramerisation domain thought to be critical for p53-mediated transactivation[556].

The initial specificity of Nutlin-3A for disruption of the MDM2-p53 interaction has now been questioned in several reports. Cells with mutant p53 have shown induction of apoptosis in response to Nutlin-3A treatment[220]. This appears to be through activation of the p53 homolog p73 via direct disruption of p73-MDM2 binding or induction of p73 expression by E2F1[353,557,558]. A proteomic study has also revealed widespread dysregulation of the MDM2 interactome by Nutlin-3A[559].

Monitoring cell proliferation in real-time using the xCELLigence assay permitted the observation of cell death commitment as a decrease in cell index on drug withdrawal (Figure 3-26C, Figure 3-27C). This finding helps to explain the lack of colonies observed in the OE33 cell line after 1 μM PLK-1A treatment (Figure 3-22).

The IC_{50} values calculated from the xCELLigence data for the KYSE-30 cell line differ from those calculated with a significantly higher number of replicates using the AlamarBlue method. The absolute values calculated from the xCELLigence data may vary with subsequent repetition. The obvious change in proliferation rate induced by drug wash-off provides sufficient evidence, however, that the observed

enhanced effect of PLK-1A after drug removal is not simply an artefact of cells being washed off with the change of media (Figure 3-22, Figure 3-23).

The enhancement of sensitivity on withdrawal of PLK-1A was more marked with the oesophageal cancer cell line KYSE-30 than the non-transformed CP-A cell line. This was confirmed in OAC cell lines by the colony outgrowth assay.

The mechanism of cell death was not investigated in this study. Previous reports in ESCC have suggested PLK-1 suppression can induce apoptosis via induction of Caspase-3[269]. This could be confirmed in OAC cell lines after pulsed PLK-1 inhibitor treatment by western-blotting through the demonstration of PARP cleavage and pro-caspase-3 reduction or through flow-cytometry and demonstration of a sub G0/G1 peak[560]. Caspase-3 activation can also be demonstrated using a specific luminescence assay[561] or even in live cells using a fluorescence biosensor[562].

Minimal changes in sensitivity were observed after wash-off of BI2536 in the xCELLigence assay and colony outgrowth assay. Both compounds inhibit kinase activity by competing for ATP-binding; although PLK-1A is reportedly more selective for PLK-1 over PLK-2 (Table 3-5).

Early phase clinical trials of BI2536 demonstrate dose limiting haematological toxicity and tumour response rates have been limited. A mechanism to explain the relative non-selective *in vivo* cytotoxicity of BI2536 can be proposed on the basis of the recent publication of a new PLK-1 inhibitor, SBE13, that traps the protein in an inactive conformation and has exclusive specificity towards PLK-1[563]. Testing of SBE13 on various non-transformed cells *in vitro* produced no cytotoxicity and only transiently slowed cell progression in stark contrast to BI2536[563] . This lends weight to the hypothesis that drugs with higher specificity for PLK-1 may have preferential *in vivo* phenotypes.

A further possibility lies in the drug half-life. The sensitivity to PLK-1A was shown to be enhanced by pulsed administration as a 72 hour (Figure 3-22), 48 hour (Figure 3-26, Figure 3-27), 24 hour (Figure 3-23 lower figures) or a 16 hour treatment (Figure 3-29) compared to 96 hour continuous exposure (Figure 3-20). This effect

appears to be specific to PLK-1A over BI2536 and may be due to the estimated 10-fold shorter half-life.

In conclusion, methods to enhance the specificity of kinase inhibitors may be successful enough to allow effective patient treatment with minimal toxicity but kinase inhibitors, especially those directed at cell cycle genes, are always likely to affect cycling cells to an extent. Other targets may however, exhibit more specific expression profiles. If a target could be identified that is exclusively expressed on oesophageal cancer cells and is available on the cell surface, anti-cancer cytotoxic antibodies could be developed that in theory should not produce any off-target effects. An approach to identify cancer specific protein expression using a proteomic method is described in the next chapter.

Chapter 4: Identifying OAC-Specific Protein Expression by Shotgun Proteomics

4.1 Chapter Aims

a) To evaluate if quantitative proteomic data can be obtained from snap frozen oesophageal adenocarcinoma and matched normal oesophageal tissue.

b) To evaluate the technical and biological reproducibility of the method.

c) To evaluate the quantitative accuracy of the method by comparing expression data with immunohistochemistry from the same tissues.

d) To compare the relative expression of quantified proteins between patient-matched tissues with published work to establish if known disease markers can be identified.

4.2 Introduction

The previous chapter built on published work identifying PLK-1 as a potential synthetic lethal candidate with mutant p53 and established PLK-1 as a candidate therapeutic target in OAC. Despite the encouraging data with the short half-life compound, systemic toxicity is still a significant concern with PLK-1 inhibitors due to the ubiquitous role of this kinase in the cell cycle. Narrowing the range of tissues targeted could enhance the therapeutic index.

Cytotoxic antibodies directed against tumour-specific proteins provide potentially highly specific therapies. This chapter will investigate the use of quantitative proteomics to identify tumour-specific proteins from human oesophageal tissue samples. By measuring protein expression directly, tissue-specific changes in protein abundance from post-transcriptional mechanisms are accounted for, providing an advantage over transcriptome-based approaches.

Changes to the extracellular microenvironment can lead to significant changes in tumour cell phenotype, potentially via relative changes in protein abundance[360]. Human tissue samples were therefore used rather than oesophageal cell lines to attempt to capture the *in vivo* expression profile of OAC cells. As pre-analytic factors, such as warm ischaemic time, can have a significant bearing on the quantitative data obtained from proteomic studies, a high-quality tissue resource was established for this study[564]. Tissue samples were collected from prospectively-consented patients undergoing resection for OAC and snap –frozen.

4.2.1 Quantitative Proteomics

Accurate relative peptide quantitation can be achieved in shotgun proteomics by the use of isobaric peptide labels. For this study, Tandem Mass Tags (TMT, Proteome Sciences) were used allowing up to six different samples to be compared simultaneously[430]. To reduce sample complexity and enhance the number of confident protein identifications, tryptic peptide mixtures were fractionated. An orthogonal combination of fractionation by isoelectric point using OFFGEL electrophoresis and retention time on a reverse phase chromatographic column was utilised[565].

4.3 Results

4.3.1 Tissue samples

Tissues from two clinically-matched patients differing only significantly in the number of lymph node metastases were used for this study. The clinical characteristics of each patient are summarised in Table 4-1. TMT-labelled tryptic peptides from snap-frozen OAC and patient-matched normal squamous epithelium from both patients were pooled, fractionated by OFFGEL electrophoresis and then each of the 12 fractions analysed three times by LC-MS/MS. Mass spectra from all experiments were pooled and analysed together using the Easyprot platform[504].

Table 4-1 Clinical Characteristics of Patients Donating Samples for this Study

	Patient	
	15	17
Gender	Male	Male
Age	63	52
Comorbidities	Nil	Nil
Histology	Adenocarcinoma	Adenocarcinoma
Location	OGJ Siewert Type I	OGJ Siewert Type I
Neoadjuvant Therapy	2xCisplatin, 5-Fluorouracil (2xCF)	1xCF (Abandoned due to Acute Left Leg Ischaemia)
Post-Chemotherapy Radiological Response Evaluation (CT)	Stable Disease	Stable Disease
Surgery	Ivor-Lewis Oesophagectomy	Ivor-Lewis Oesophagectomy
Maximum Tumour Diameter	45 mm	43 mm
Distance to Proximal and Distal Margins	>1 mm	>1 mm
Distance to Circumferential Margin	0.2 mm	0.4 mm
Differentiation	Poor	Poor
Histological Lymphovascular Invasion	Yes	Yes
Histological Venous Invasion	Yes	Yes
Histological Perineural Invasion	Yes	Yes
Resection Complete	R1	R1
T stage	T3	T3
Number of Positive Nodes	17	0
Total Number Nodes Resected	29	19
AJCC Stage	IIIC	IIB
Alive at Time of Analysis	No	Yes
Overall Survival From Diagnosis	19.8 months	27.4 months (censored)

AJCC – American Joint Committee on Cancer, OGJ – Oesophagogastric Junctional Tumour, R1- Resection Microscopically Incomplete, CT – Computed Tomography.

4.3.2 Peptide and Protein Quantitation

A total of 1,097 proteins were identified with a peptide FDR<1% and two unique peptides per protein. The mean peptide reporter ratios between OAC and normal oesophagus (TvO) for each protein, for each patient were calculated and the ratios normalized so the median $\log_2$ ratio = 0. The normalized distributions are shown in Figure 4-1

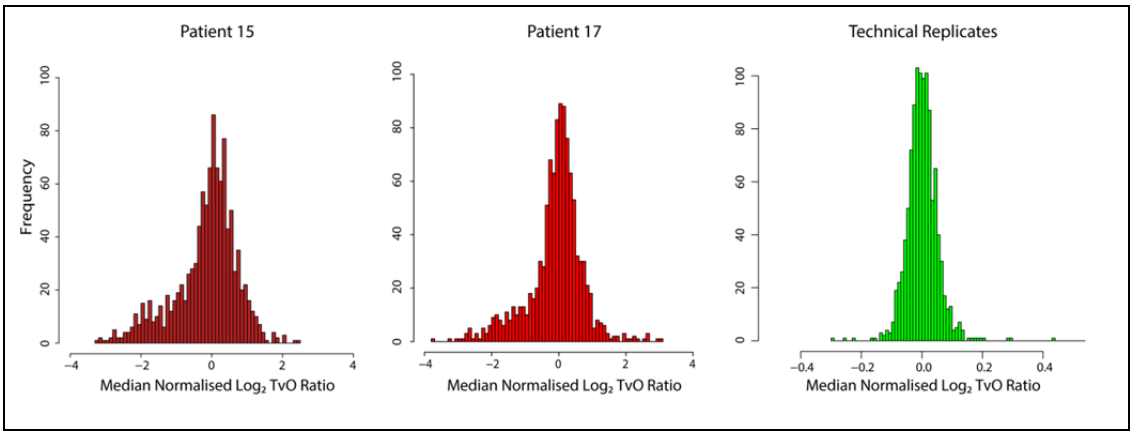

Figure 4-1 Histograms of Median Normalised Protein Ratios for each Patient and for Technical Replicates. Note the differing x-axis scales.

The protein ratios for each patient and for the technical replicate showed approximately Gaussian distributions but the variances differed significantly (Fligner-Killeen p<0.05). The TvO ratios were therefore compared to the technical replicate ratios using Welch's statistic. Welch's modified t-test is a parametric statistic that can be applied to compare two means from approximately normally distributed populations with unequal variances to test the hypothesis that the means are the same[512]. This statistic is also relatively robust to deviation from normality when the sample size is large. For patient 15, 526 proteins and for patient 17, 529 proteins were significantly dysregulated (FDR-p <0.05) with 430 proteins dysregulated in both patients. The distribution of proteins by TvO ratio and FDR-corrected p-value is displayed for each patient in Figure 4-2 (see Methods 2.11). This highlights, as expected, that the majority of proteins with a p>0.05 have an unchanged expression between OAC and normal oesophagus or were only quantified by a few peptides.

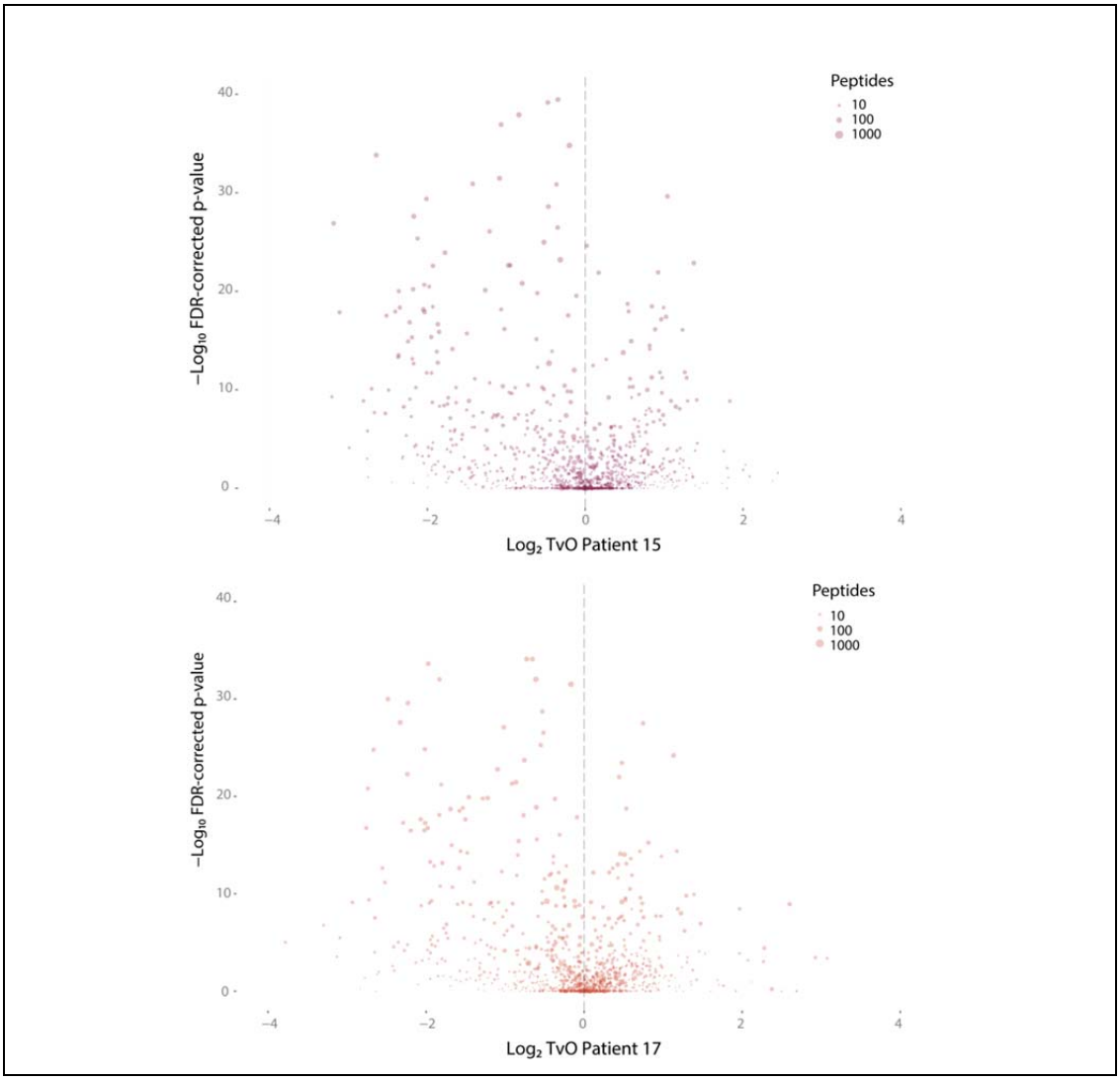

Figure 4-2 Volcano Plots Illustrating the Distribution of Proteins According to TvO ratio and FDR-corrected p-value. Points represent proteins and are scaled according to the number of peptides quantified. FDR correction was performed with the Benjamini-Yekutieli method[566].

4.3.3 Comparison of Protein Expression Between Patients

The relationship between the TvO ratios from each patient was next examined. The correlation in TvO ratios between patients was good (Figure 4-3, Pearson Correlation Coefficient = 0.6726, $p<0.001$) which is reassuring given the close clinical match for the patients. The number of lymph node metastases was significantly different between patients, however, and this is the clinical variable with the strongest association with prognosis[117]. Proteins differentially expressed in primary tumours associated with extensive lymph node metastases may be associated with targetable pathways with a role in metastasis. The outliers were therefore annotated for further consideration (Figure 4-4).

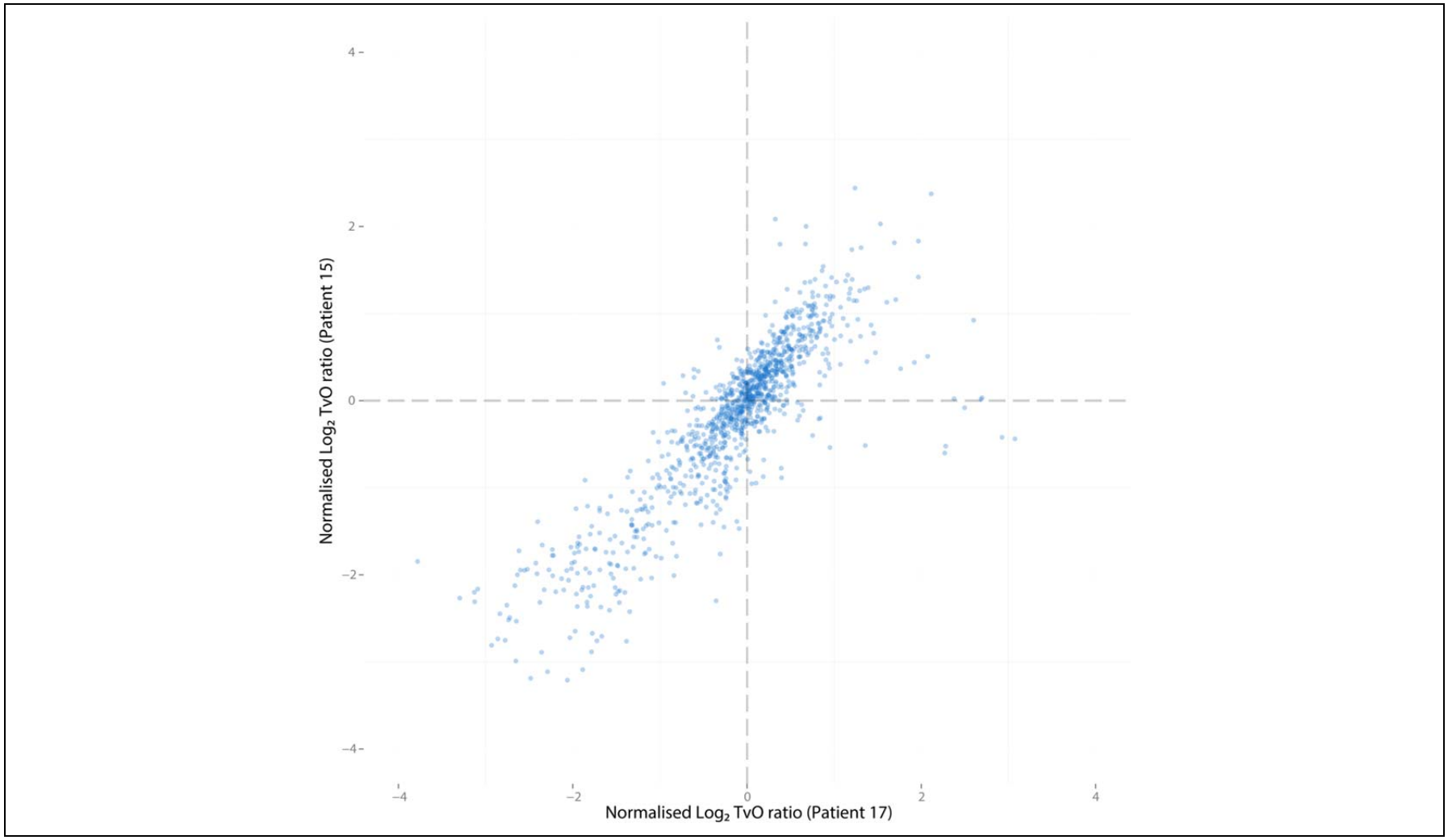

Figure 4-3 Comparison of TvO ratios from Patient 15 and Patient 17. Each point represents a protein. The position on the plot is determined by the Log_2 TvO expression ratio for each patient.

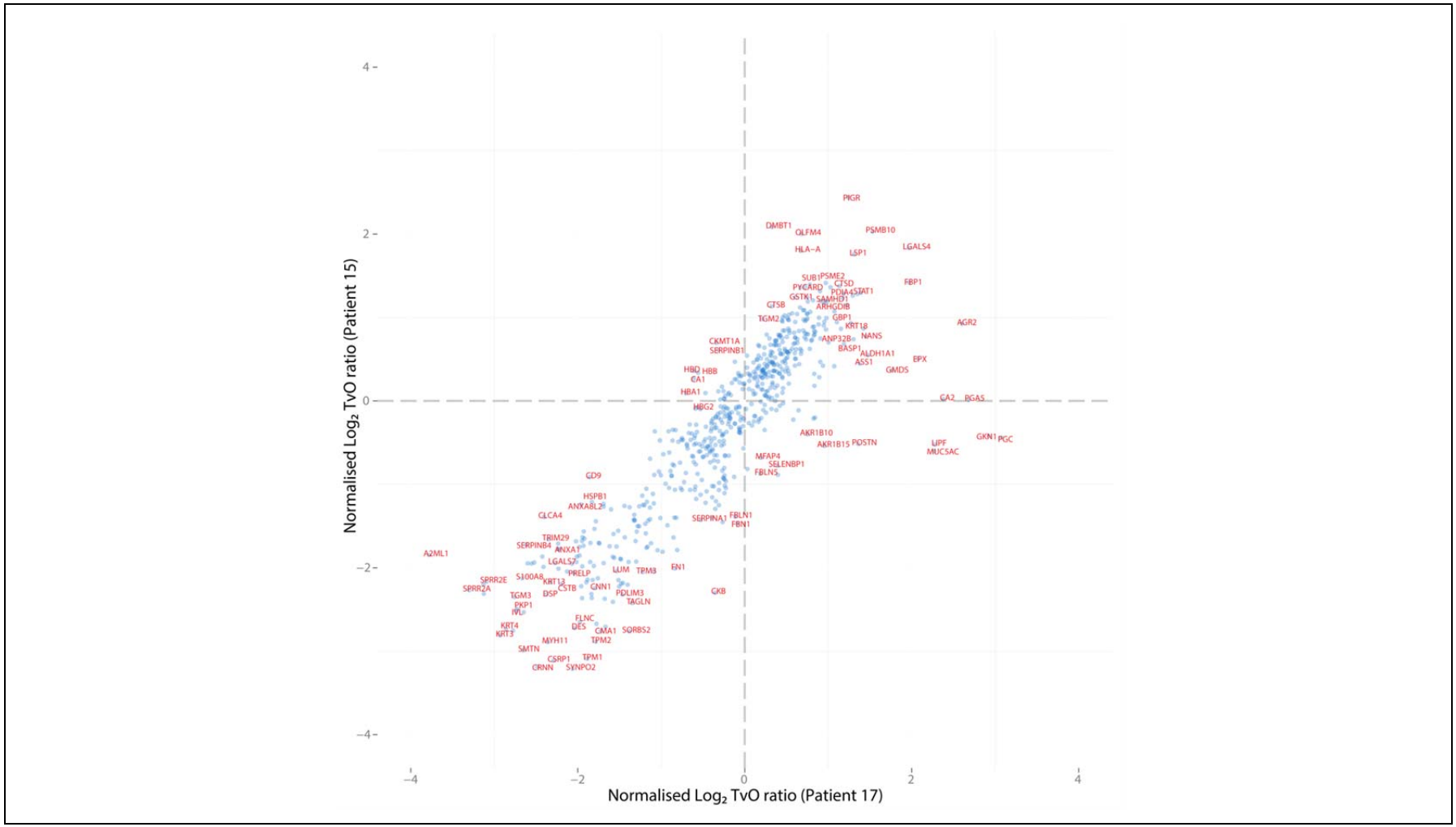

Figure 4-4 Comparison of Proteins with Significantly Dysregulated TvO Ratios. Each point represents a protein. The position on the plot is determined by the Log_2 TvO expression ratio for each patient. Only proteins with an FDR-p<0.05 are displayed and selected outliers have been annotated with gene names.

4.3.4 Evaluation of Significantly Dysregulated Proteins

Proteins differentially up or down-regulated in patient 15 but relatively unchanged in patient 17 were considered. Although no definitive conclusions can be drawn from this comparison, proteins involved in extracellular matrix homeostasis were dysregulated between patients including Cathepsin B (CTSB), Fibulin-1 (FBLN1), Fibulin-5 (FBLN5), Fibrillin-1 (FBN1) and Fibronectin (FN-1). Matrix remodelling is a common feature during cancer cell invasion and metastasis and the down-regulation of FBLN1 and FBLN5 has been associated with oncogenesis in prostate, gastric and liver tissue[567-569]. CTSB, a cysteine protease, has also been implicated in cancer invasion and metastasis and may be a target for cancer-specific imaging probes[570-573].

This suggests the comparative proteomic approach may yield insight into the biology of OAC and uncover cancer-specific drug targets. To examine how the quantitative results from this study compare with the published literature on OAC, the significantly dysregulated proteins (FDR-corrected $p < 0.05$ in either p15 and p17) were ranked by the sum of TvO ratios and the top 150 proteins upregulated in both p15 and p17 considered in greater detail. A selection of these proteins along with a summary of the evidence for their upregulation in OAC or cancer in general is provided in Table 4-2. A smaller selection of proteins down-regulated in both patients was also examined. The complete list of significantly dysregulated proteins and the accompanying TvO ratios, technical replicate ratios, FDR-corrected p-values and number of peptides quantified is presented in the Appendix.

4.3.5 Validation of TMT Quantitative Accuracy

This comparative proteomic approach relies on the quantitative accuracy of the isobaric TMT ratios. To test this, six proteins predicted to be upregulated and two proteins predicted to be down-regulated by varying degrees in both patients' tumours were selected for validation by immunohistochemistry (IHC). These proteins are highlighted in purple in Table 4-2 and their relative expression in comparison to other significantly dysregulated proteins demonstrated in Figure 4-5.

Table 4-2 Selected Significantly Dysregulated Proteins. Examples of the evidence either from hypothesis directed research or published microarray data supporting the expression profile are provided in the corresponding references. Candidates selected for IHC are coloured purple.

Upregulated Proteins						
Protein Name	Patient 15 TvO Ratio	Patient 15 FDR p-value	Patient 17 TvO Ratio	Patient 17 FDR p-value	Evidence of Dysregulation in OAC	Evidence of Dysregulation in other Cancers
Galectin-4	3.56	1.40E-09	3.91	3.54E-09	Yes[60,61,277]	
Polymeric immunoglobulin receptor	5.43	0.026467	2.36	0.00366	Yes[574]	
Anterior gradient protein 2 homolog (AGR2)	1.90	1.19E-22	6.06	1.15E-09	Yes[60,61,171,277]	Yes[575]
Fructose-1,6-bisphosphatase 1	2.67	2.61E-05	3.91	0.00013	Yes[60,61,277]	Yes (down-regulated)[576,577]
Lymphocyte-specific protein 1	3.38	0.000174	2.48	0.000331		Yes[578]
Coactosin-like protein	2.46	1.37E-09	2.62	1.18E-10		Yes[579]
Olfactomedin-4	4.01	0.013774	1.60	0.019976	Yes[60,61,277]	
Gastricsin	0.74	3.05E-09	8.42	0.000431		Yes[580]
Signal transducer and activator of transcription 1-alpha/beta	2.40	1.63E-12	2.45	1.62E-10	Yes[60,61,171,277]	Yes[581]
Cathepsin D	2.60	1.35E-23	2.19	6.32E-25	Yes[582]	
Coronin-1A	2.44	6.06E-12	2.27	3.69E-09		Yes[583]
Deleted in malignant brain tumours 1 protein	4.24	0.06183	1.25	0.005391	Yes[584]	
Protein disulphide-isomerase A4	2.35	8.18E-17	2.25	4.16E-15	Yes [60,61,171,277]	
Keratin, type II cytoskeletal 8	2.22	5.19E-09	2.34	1.02E-08	Yes [60,61,171,277]	
Keratin, type I cytoskeletal 18	1.91	1.32E-08	2.41	6.39E-07	Yes[585]	
Cathepsin Z	2.29	0.003016	1.98	0.004004	Yes [60,61,171,277]	
Deoxynucleoside triphosphate triphosphohydrolase (SAMHD1)	2.30	8.44E-05	1.94	5.66E-06	Yes [171]	
Rho GDP-dissociation inhibitor 2 (D4GDI)	2.24	0.000149	1.93	0.000382	Yes [447]	
Thymidine phosphorylase	2.29	8.59E-09	1.88	1.98E-08	Yes[586]	
Interferon-induced guanylate-binding protein 1	1.92	0.006879	2.15	0.024065		Yes[587]
Proliferating cell nuclear antigen	1.67	0.000113	2.47	0.001047	Yes[588]	
Retinal dehydrogenase 1	1.46	1.11E-18	2.78	1.20E-07	Yes [449]	
Argininosuccinate synthase	1.36	0.00011	2.59	0.00704	Yes[589]	
Proteasome activator complex subunit 1	2.13	1.34E-09	1.64	2.61E-10	Yes[448]	

Continued overleaf….

Protein Name	Patient 15 TvO Ratio	Patient 15 FDR p-value	Patient 17 TvO Ratio	Patient 17 FDR p-value	Evidence of Dysregulation in OAC	Evidence of Dysregulation in other Cancers
High mobility group protein B1	2.12	0.00205	1.63	0.010973		Yes[590]
Protein disulphide-isomerase A3	2.06	2.31E-30	1.68	3.13E-28		Yes[591,592]
Plastin-2	1.98	3.66E-08	1.68	2.67E-08		Yes[593]
Endoplasmin	1.85	7.29E-17	1.76	5.72E-16	Yes[594]	
Annexin A4	2.06	0.060022	1.54	0.014045	Yes[60,61,169,277]	
Voltage-dependent anion-selective channel protein 1	1.99	0.000969	1.42	0.000276	Yes[277]	
Cathepsin B	2.19	0.022038	1.25	0.000199	Yes[595,596]	Yes[573]
Heterogeneous nuclear ribonucleoproteins A2/B1	1.76	7.40E-15	1.51	2.50E-14		Yes[597]
Calreticulin	1.52	4.97E-11	1.60	1.50E-07		Yes[598,599]
Tumour protein D52	1.54	0.001337	1.48	0.000751		Yes[600]
Glutathione S-transferase omega-1 (GSTO-1)	1.51	0.001498	1.45	0.002081	Yes[447,601]	Yes[602]
Inorganic pyrophosphatase	1.50	0.001059	1.42	0.024315	Yes[447]	
60 kDa heat shock protein, mitochondrial	1.50	1.16E-15	1.43	9.48E-15	Yes[447]	Yes[603]
Peptidyl-prolyl cis-trans isomerase B	1.59	0.06734	1.30	0.004522	Yes[449]	
Cofilin-1	1.47	5.25E-11	1.35	3.69E-08		Yes[604]
78 kDa glucose-regulated protein (GRP-78)	1.40	1.66E-14	1.39	6.98E-10	Yes [60,277,446,594]	
Down-regulated Proteins						
Cornulin	0.1097	1.27E-27	0.17886	1.03E-30	Yes[605]	
Keratin, type II cytoskeletal 4	0.150119	1.48E-43	0.137715	1.20E-42	Yes[606]	
Involucrin	0.174528	3.06E-18	0.150044	1.57E-21	Yes[607,608]	
Transglutaminase 3 (TGM-3)	0.196284	4.51E-19	0.147945	1.75E-17	Yes[609]	
Protein S100-A8	0.229294	4.39E-26	0.157754	1.63E-25	Yes[610]	
Small Proline Rich Protein 3	0.297078	0.00010673	0.253345	6.09E-05	Yes[60,61,277]	
Heat shock protein beta-1 (HSP-27)	0.431919	8.39E-27	0.281171	9.63E-33	Yes[625]	

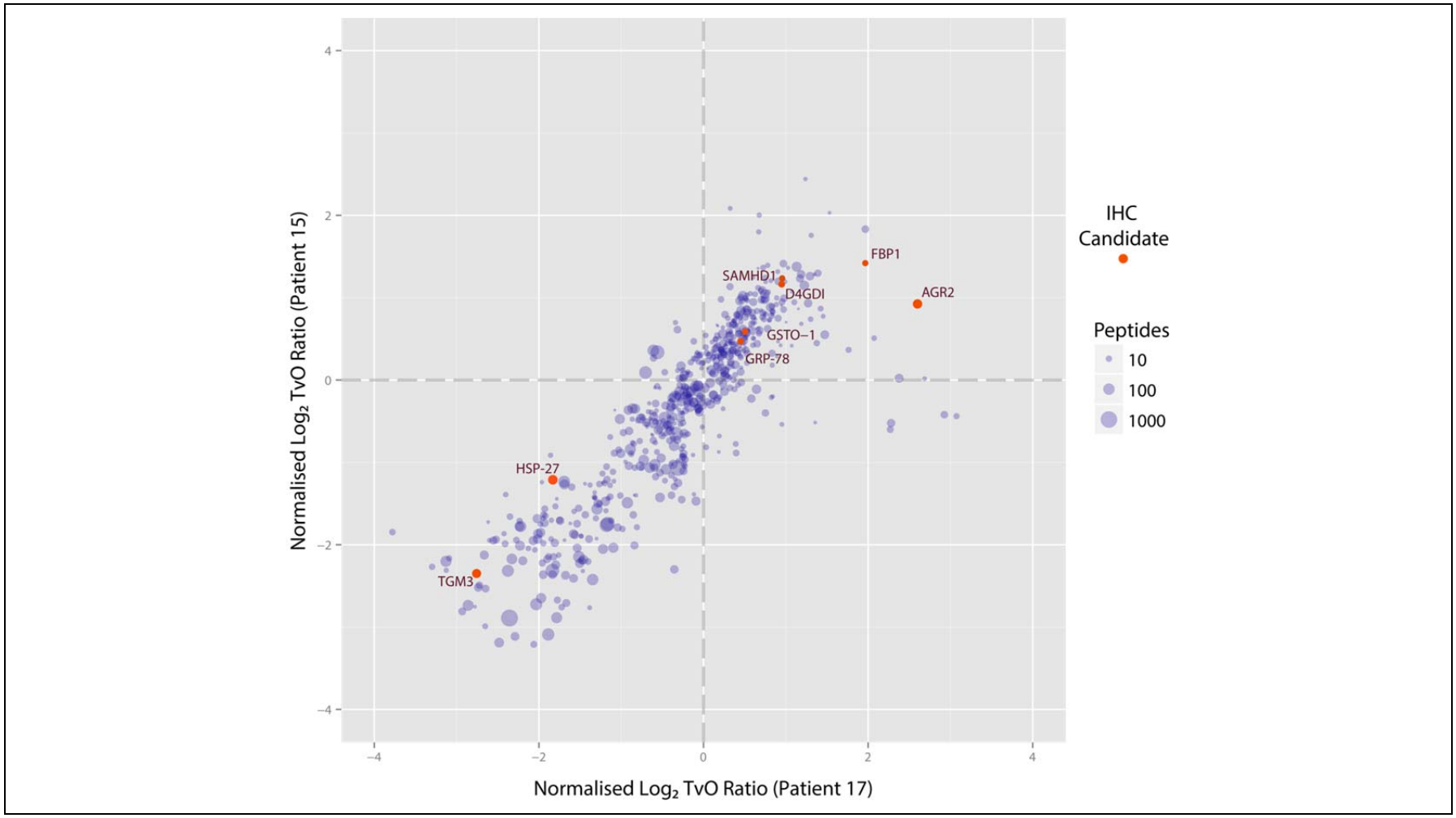

Figure 4-5 Comparison of Significantly Dysregulated Protein TvO Ratios. Each point represents a protein. The position on the plot is determined by the Log_2 TvO expression ratio for each patient. Only proteins with an FDR-p<0.05 are displayed and proteins selected for IHC are coloured red and annotated with their gene names. Points have been scaled according to the number of quantified peptides.

Confirming the specificity of antibodies used for IHC is important to prevent false positives. Antibodies predicted to be suitable for both western blotting and IHC were sought for each of the 8 candidate proteins. A panel of oesophageal cells lines incorporating the various stages of oesophageal carcinogenesis were then used to evaluate the relative expression of each candidate (Figure 4-6A).

Bands of the predicted molecular weight were detected for each candidate confirming the sensitivity of the primary antibodies. Similar to the tissue expression profiles, there was generally higher expression in the adenocarcinoma cell lines for AGR2, D4GDI and SAMHD1 compared to primary squamous and Barrett's cells although some heterogeneity was observed. In contrast both HSP-27 and TGM3 were expressed at higher levels in primary squamous cells and a doublet band was observed specifically in squamous cells with the TGM3 antibody suggesting differential regulation of the protein between these cells and the other cell lines. Ubiquitous expression of varying levels was observed for GSTO-1, GRP-78 and HSP-27 consistent with their roles in redox homeostasis or as chaperones.

To confirm the specificity of the antibodies to proteins predicted to be upregulated in OAC, siRNA was transfected into cell lines expressing the relevant genes. Western blots demonstrated a significant reduction in the band intensity at the predicted molecular weight for SAMHD1, AGR2, D4GDI and GRP78 when siRNA targeted to the corresponding gene was transfected. This confirmed the specificity of the antibodies towards these targets. A minimal reduction was observed with siRNA to GSTO-1. This may be due to poor mRNA targeting of the transfected siRNA sequence, a long protein half-life, non-transcriptional control of protein abundance, or poor antibody specificity.

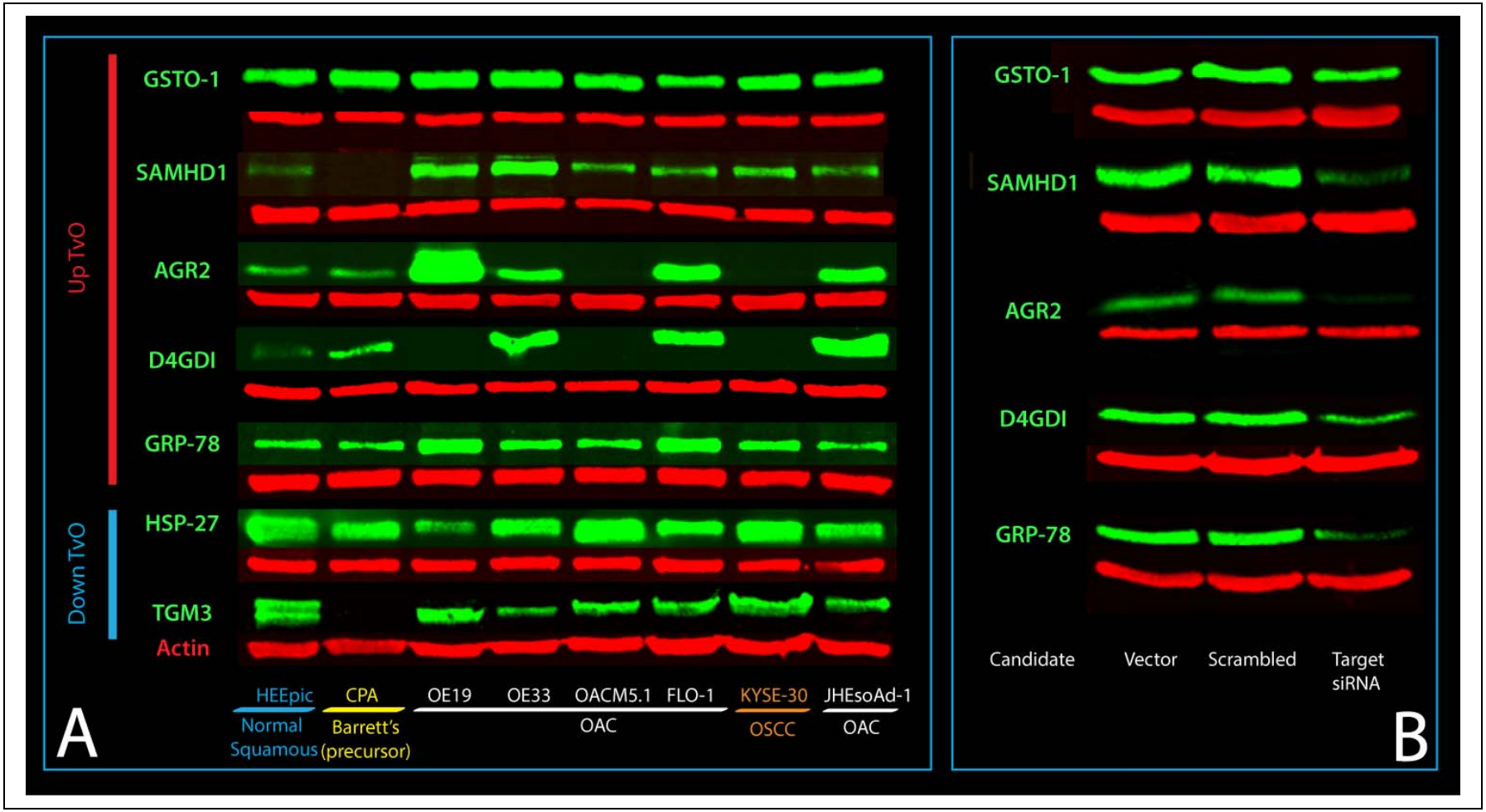

Figure 4-6 Expression of Candidates in a Panel of Oesophageal cells and Validation of Antibodies for IHC. (A) Lysates from each of the representative cell types growing under basal conditions were resolved by SDS-PAGE (20µg per lane using 8-12% gels) and western blots probed with the indicated primary antibodies before scanning using a Li-Cor imaging system. (B) KYSE-30 cells (GRP-78), OE33 cells (D4GDI, AGR2, SAMHD1) or CPA cells (GSTO-1) were transfected with siRNA or vector alone and harvested at 72 hours post-transfection. Lysates were then resolved by SDS-PAGE and western blotting carried out as previously.

The evaluation of FBP-1 expression was more difficult as two initial antibodies detected multiple bands at higher molecular weights but none at the predicted weight of 37 kDA (data not shown). A third antibody (Atlas Antibodies, HPA005857) detected a band of the appropriate molecular weight amongst multiple others and this band was highly expressed in the breast cancer cell line MCF7 (Figure 4-7 B), previously published as expressing FBP1[611]. When MCF7 cells were transfected with siRNA to FBP1, there was a reduction in the intensity of the 37 kDa band confirming the specificity of this antibody (Figure 4-7 A). This gene has been shown to be overexpressed in OAC compared to normal oesophagus at the mRNA level (Table 4-2)[61] and therefore qRT-PCR was performed confirming the results of western blotting (Figure 4-7C).

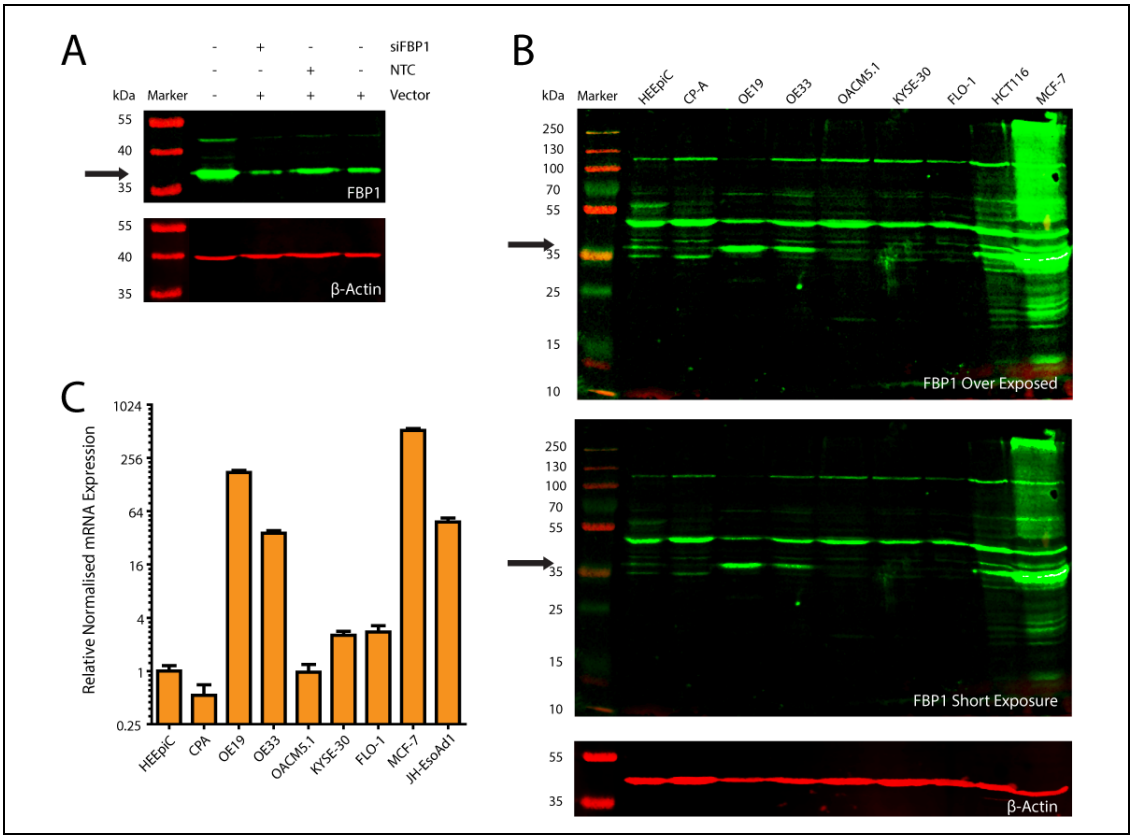

Figure 4-7 Evaluation of FBP1 Expression in a Panel of Oesophageal Cell Lines. (A) Western blot of FBP1 expression in MCF7 cells transfected with siRNA or vector alone and harvested at 72 hours post-transfection (30µg/lane). The arrow represents the predicted molecular weight of FBP1. (B) Expression of FBP1 in a panel of cell lines. Lysates from each of the representative cell types growing under basal conditions were resolved by SDS-PAGE (20µg per lane using a 10% gel) and western blots probed with the FBP1 antibody before Li-Cor scanning at high amplification (Over Exposed) or normal amplification (Short Exposure). (C) Relative FBP1 mRNA expression in a panel of cell lines grown under basal conditions. Error bars represent the SEM of 4 technical replicates from one biological replicate. Note the Y-axis is on a log scale.

Six candidates; AGR2, GRP-78, SAMHD1, D4GDI, TGM3 and HSP-27, were selected for validation of relative expression by IHC. After optimization of antigen retrieval conditions and primary antibody dilutions, sections from FFPE tissue blocks of the original resection specimens used to derive the fresh tissue samples for proteomic analysis were stained (Figure 4-8 - Figure 4-14).

All four of the proteins predicted to be upregulated in OAC compared to normal oesophagus showed higher expression in the tumour sections compared to squamous epithelium. Similarly the proteins with TvO ratios less than 1; HSP-27 and TGM3, showed the highest expression in squamous epithelium confirming the quantitative proteomic accuracy (Figure 4-12, Figure 4-13).

AGR2 was expressed in OAC and gastric epithelial cells with no evidence of expression in normal squamous cells (Figure 4-8). GRP-78 was also overexpressed in OAC cells both in the primary tumour and involved lymph nodes (Figure 4-10) compared to squamous epithelium. SAMHD1 showed more widespread nuclear expression but was upregulated in OAC cells in the primary tumour and in lymph node metastasis (Figure 4-11).

When gastric epithelial sections were examined, several staining patterns were identified. Homogeneous cytoplasmic staining was observed in all or scattered basal epithelial cells of the gastric crypts. This was a common finding with the staining characteristics associated with a false positive result[612]. This is supported by the appearance of this pattern in the "no primary" gastric slide for patient 17 (Figure 4-14). Presumed true positive gastric epithelial cytoplasmic staining was seen with AGR2 and presumed true positive nuclear staining was seen with SAMHD1 (Figure 4-8, Figure 4-11). Some DAB positivity was noted in the gastric epithelium in the slides stained for GRP-78 but this was more likely to represent false positive staining (Figure 4-10).

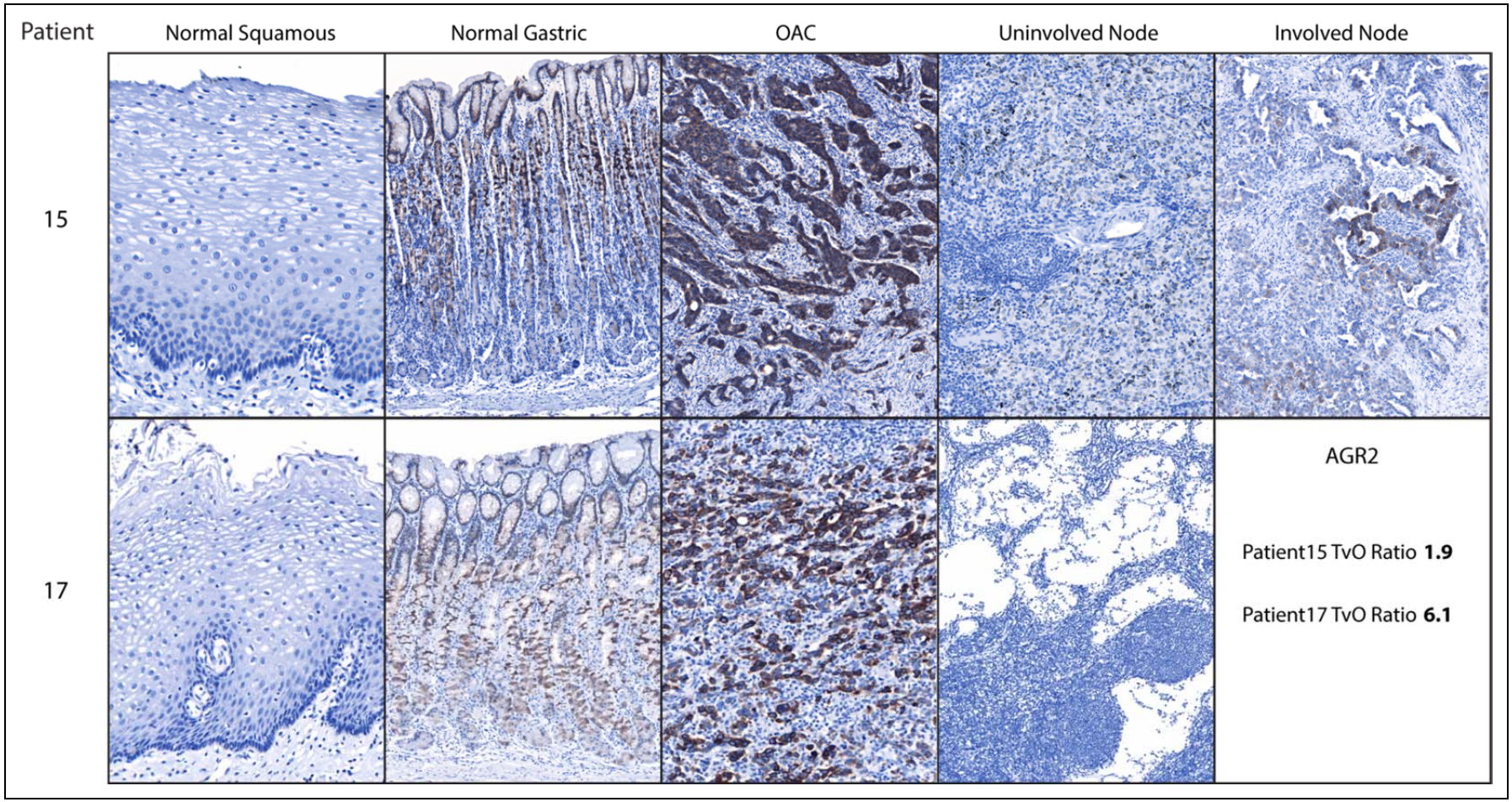

Figure 4-8 IHC staining pattern for AGR2. Representative images highlighting the staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7. The observed TvO expression ratios are provided.

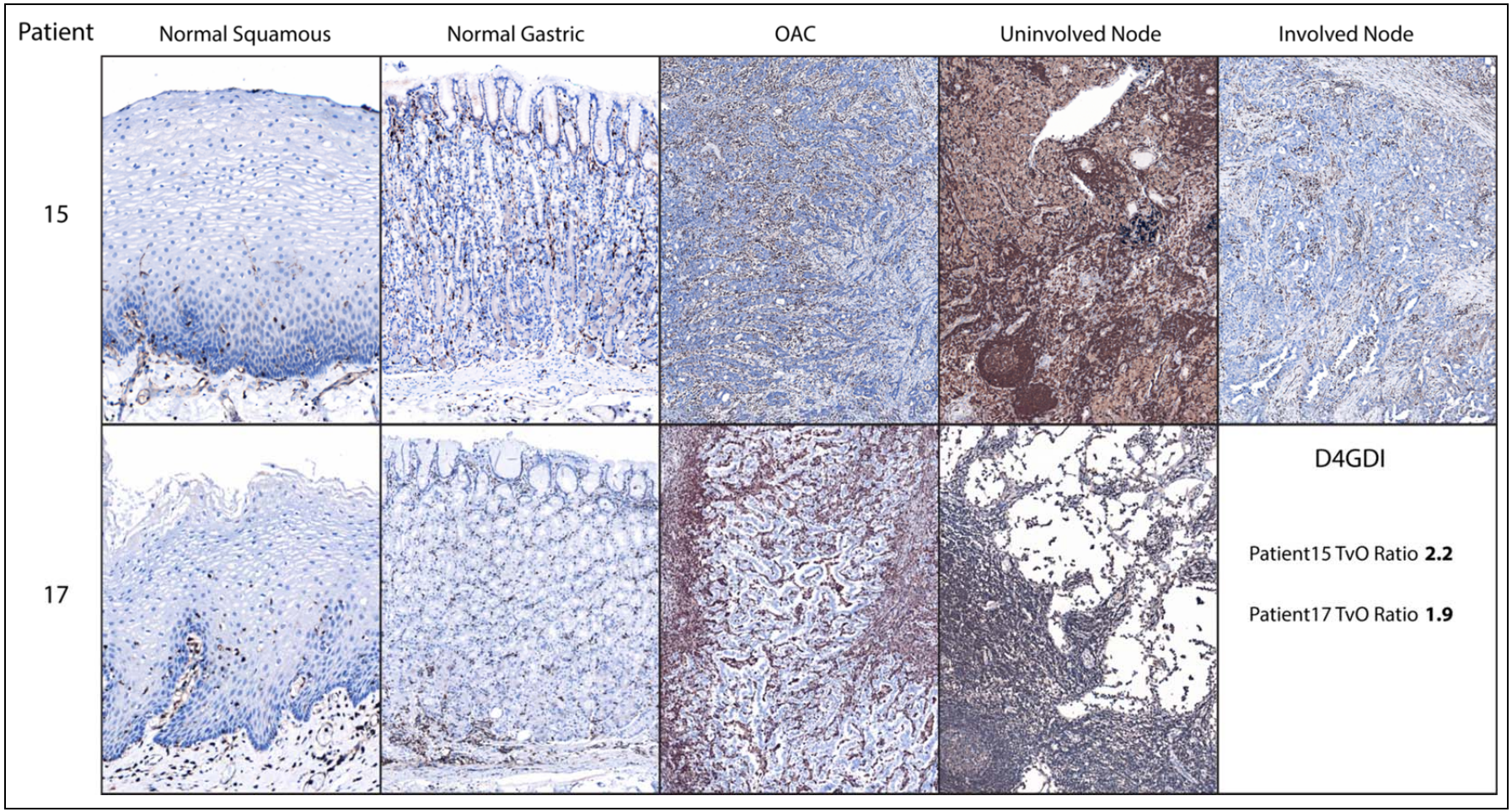

Figure 4-9 IHC staining pattern for D4GDI. Representative images highlighting the staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7. The observed TvO expression ratios are provided.

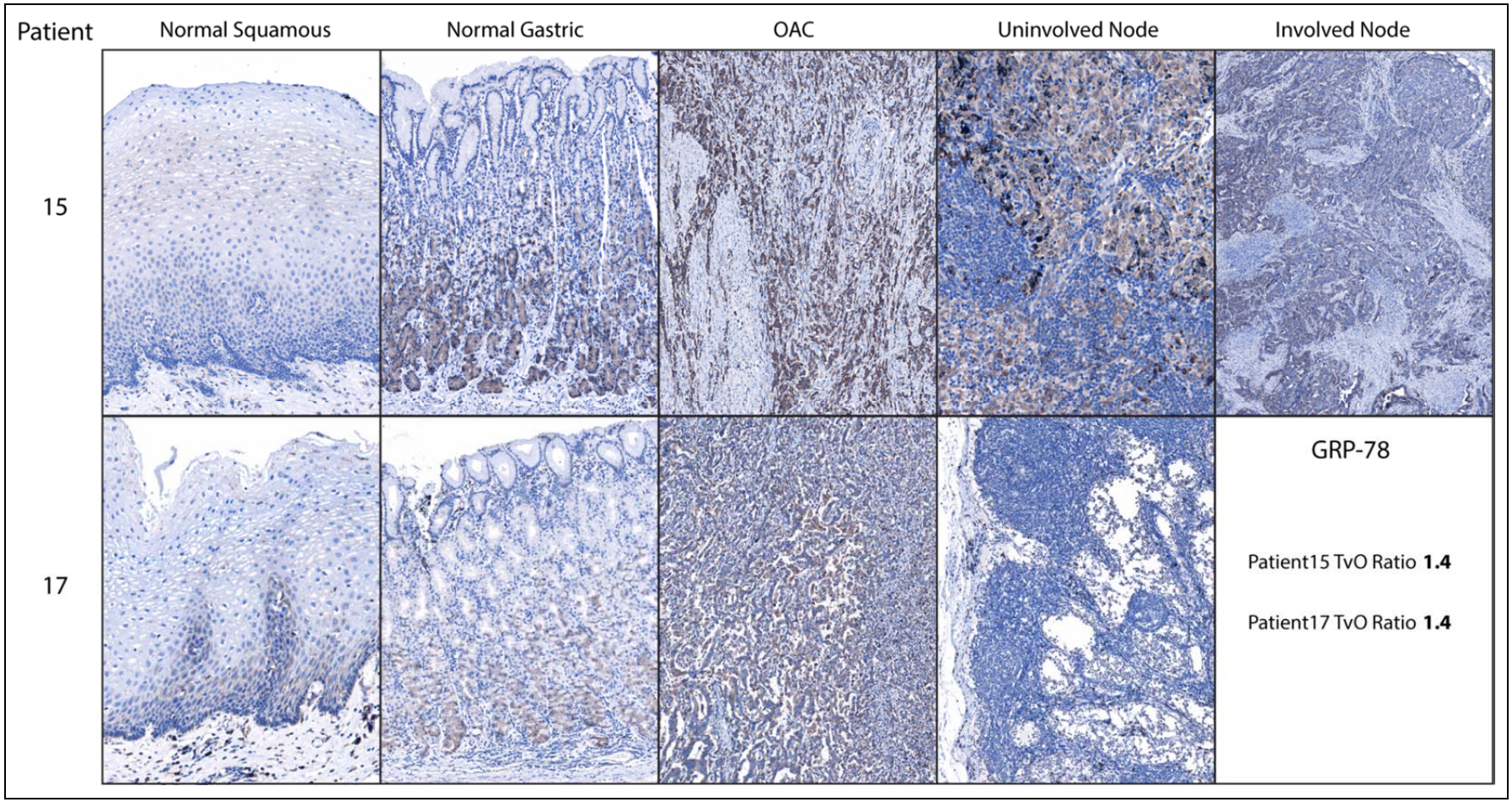

Figure 4-10 IHC staining pattern for GRP-78. Representative images highlighting the staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7. The observed TvO expression ratios are provided.

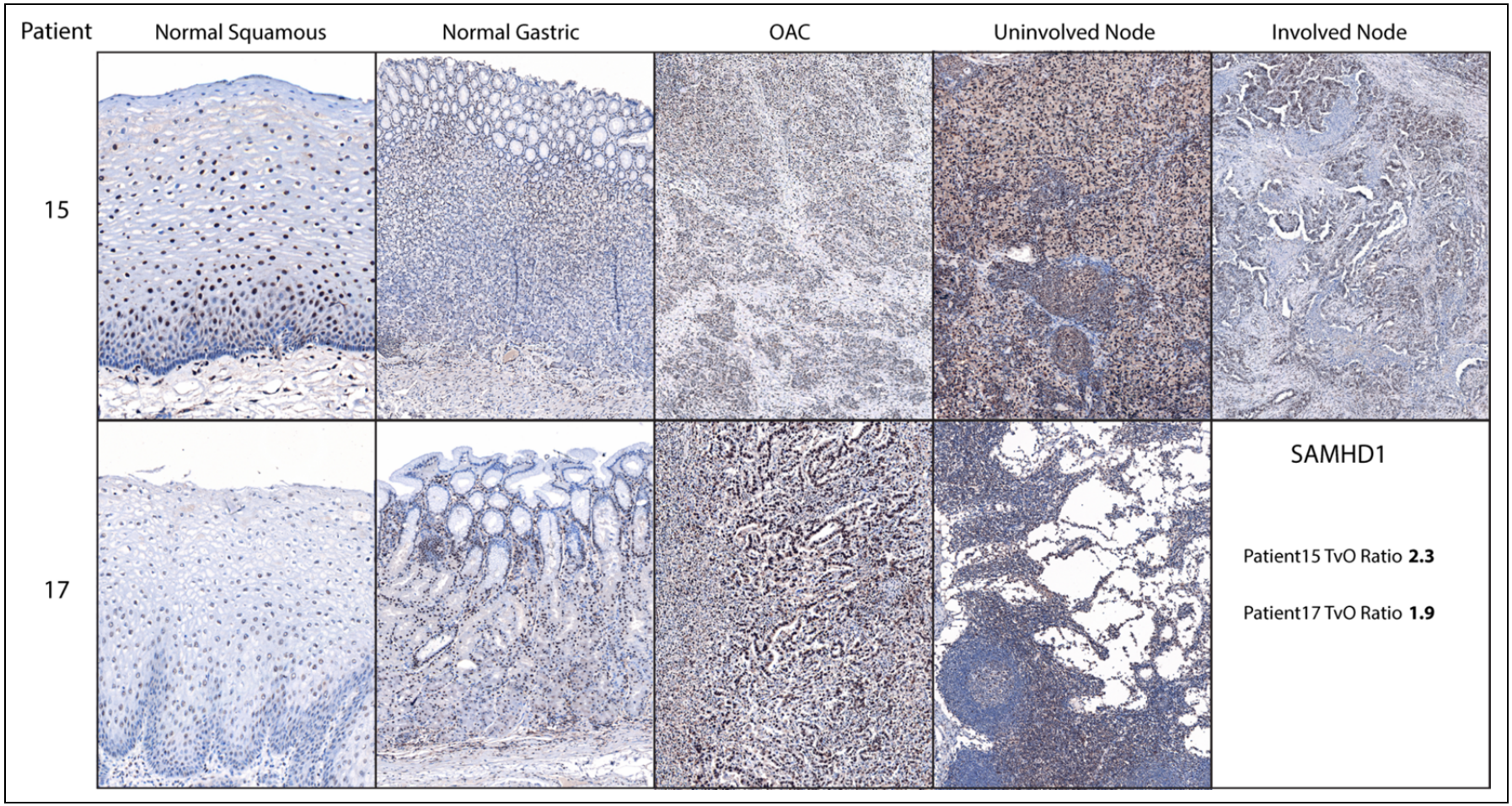

Figure 4-11 IHC staining pattern for SAMHD1. Representative images highlighting the staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7. The observed TvO expression ratios are provided.

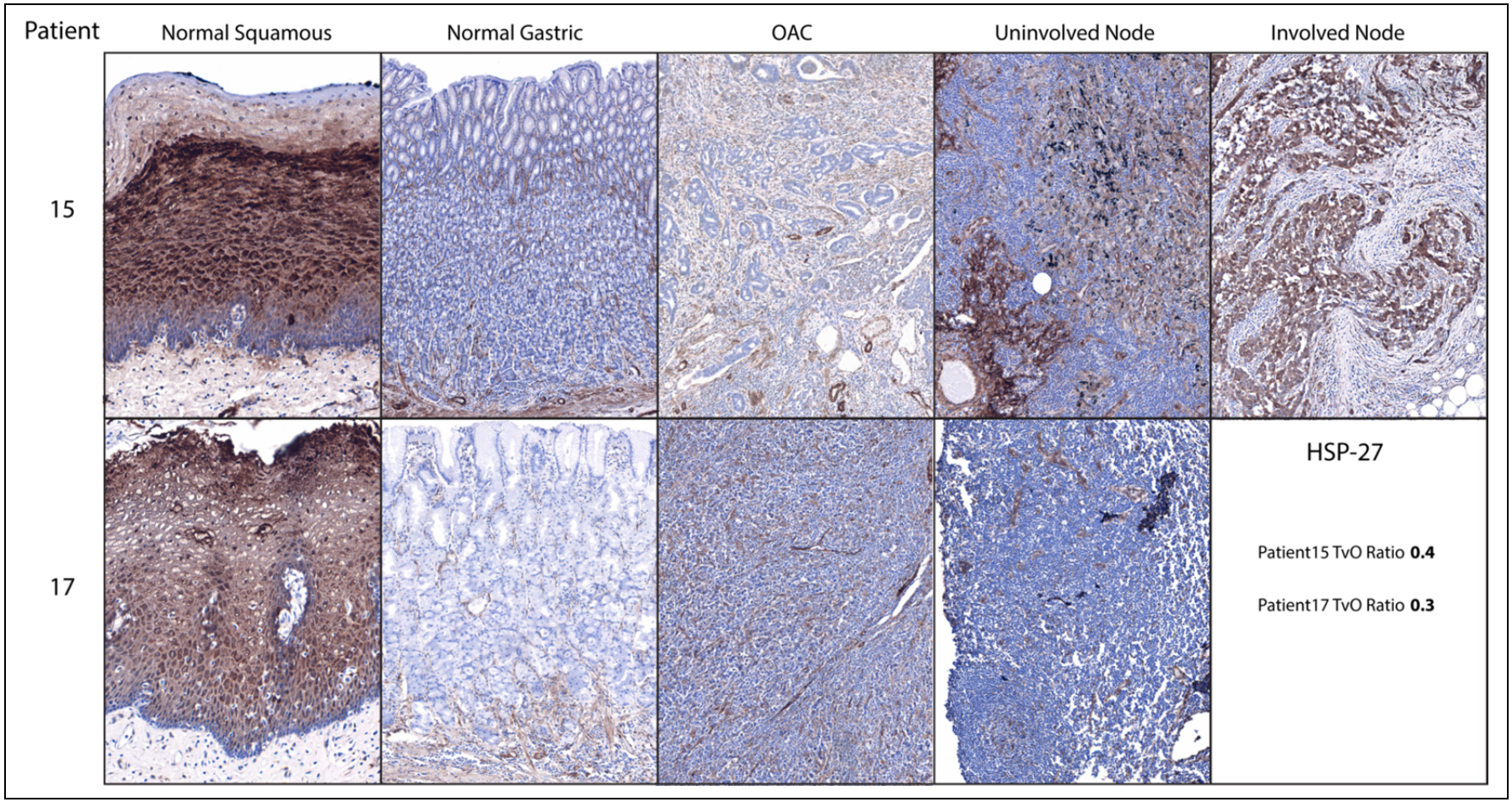

Figure 4-12 IHC staining pattern for HSP-27. Representative images highlighting the staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7. The observed TvO expression ratios are provided.

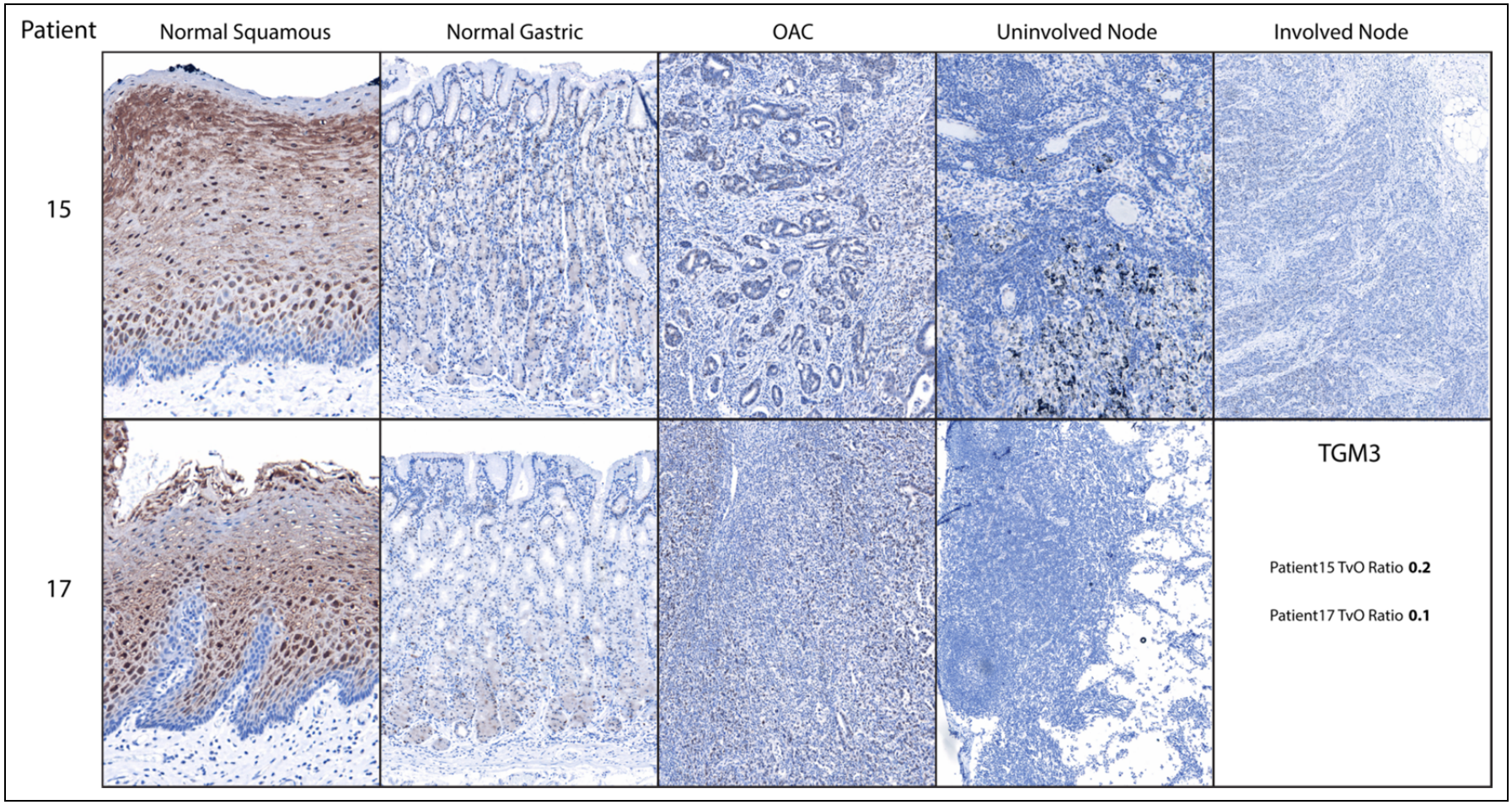

Figure 4-13 IHC staining pattern for TGM3. Representative images highlighting the staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7. The observed TvO expression ratios are provided.

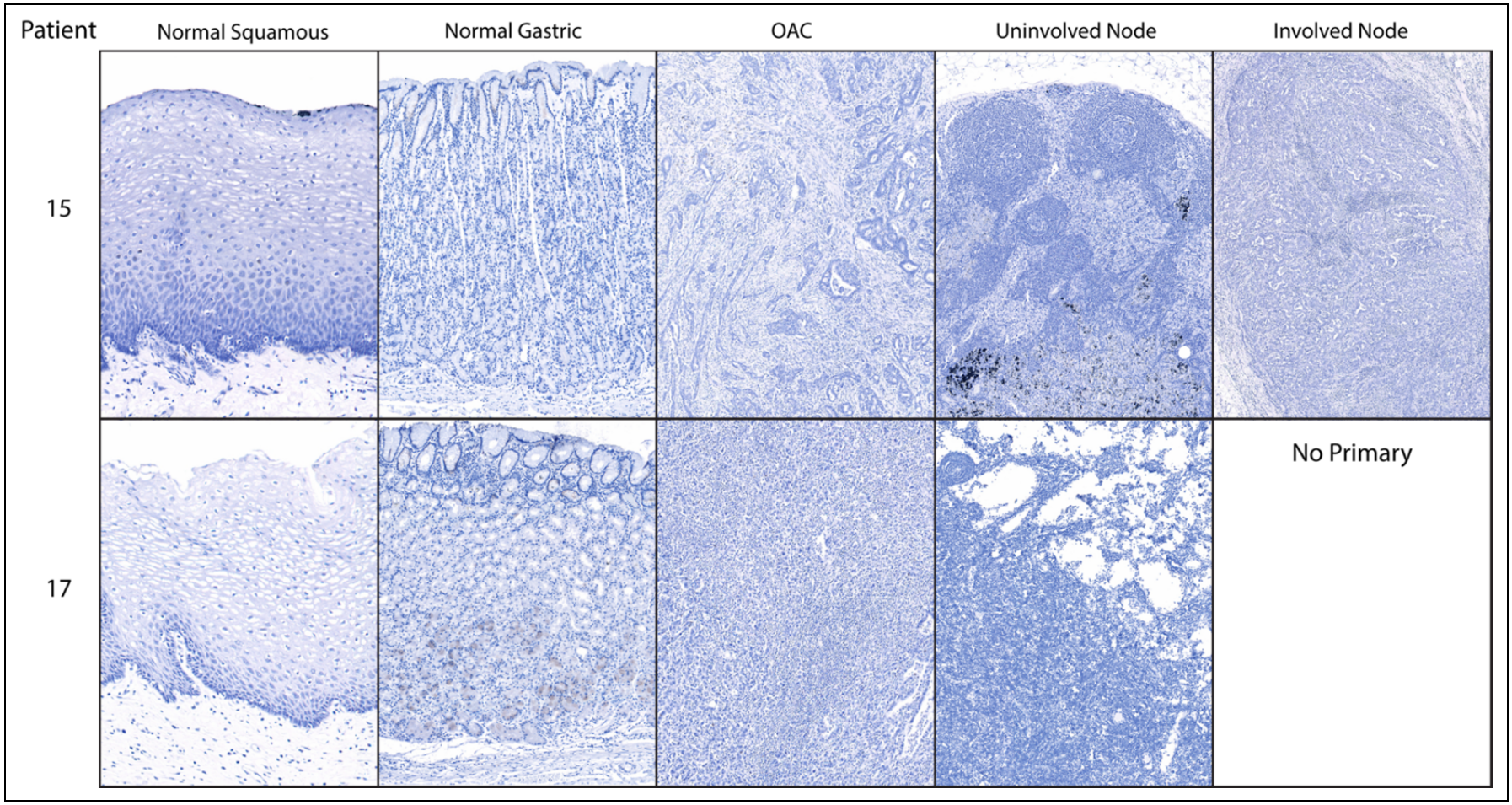

Figure 4-14 IHC staining pattern with no primary antibody. Representative images highlighting the background staining pattern from various tissues are demonstrated for each patient. Details of the staining protocol are provided in Methods 2.7.

Positive staining for D4GDI was also observed in the gastric slides (Figure 4-9). This was limited to small stromal cells typical of lymphocytes. Indeed, strong nuclear and cytoplasmic D4GDI positivity was observed in similar stromal cells in the tumour with no staining in the epithelial cells. Consistent with expression in lymphocytes, D4GDI was also highly expressed in lymphoid follicles in both uninvolved and involved nodes (Figure 4-9).

A relatively homogeneous staining pattern was observed for all the proteins except HSP-27. This protein was most highly expressed in squamous epithelium but staining across both patients' tumours was variable with some tumour cells negative and some showing high perinuclear and cytoplasmic positivity (Figure 4-15). HSP-27 positivity was also observed in muscle tissue, endothelial cells and stromal fibroblasts (Figure 4-16B). Strikingly, HSP-27 was highly over-expressed in epithelial-derived tumour cells within lymph node metastases (Figure 4-16A).

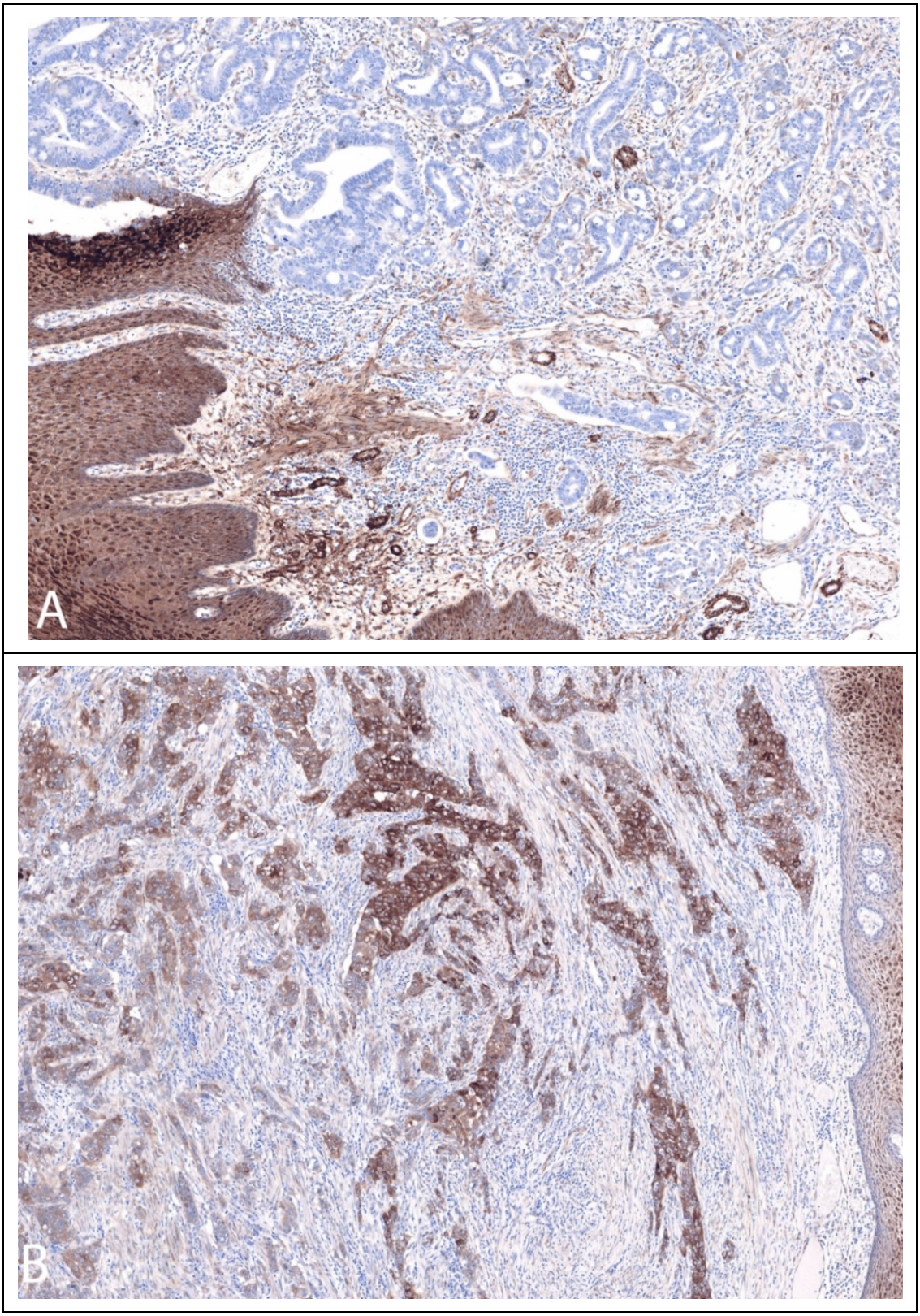

Figure 4-15 Heterogeneous Tumour Expression of HSP-27. (A) Area within patient 15's tumour with low HSP-27 expression. (B) Area within patient 15's tumour with high HSP-27 expression. Normal squamous epithelium is seen on the lower left (A) and right (B) of the images.

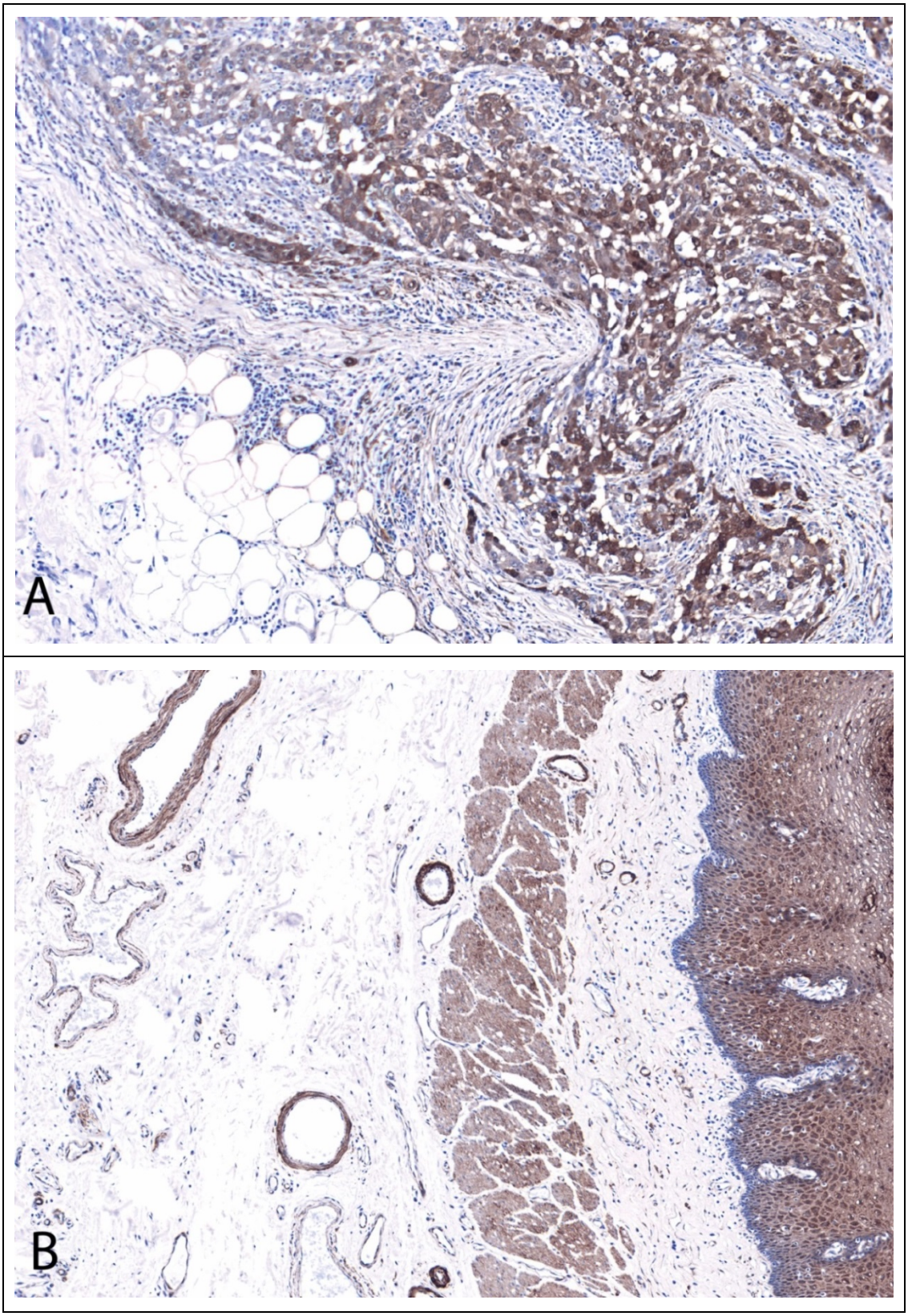

Figure 4-16 Expression of HSP-27 in a Lymph Node Metastasis from patient15 (A) and Normal Oesophageal Submucosal Tissue (B). (B) Normal squamous epithelium is on the right and Muscularis Mucosa is in the centre of the image.

4.3.6 Discussion

To the author's knowledge, this is the first study of protein expression in OAC tissue to utilize a purely shotgun proteomic approach. The confident identification and quantification of over 1,000 proteins demonstrated the potential power of this approach even when applied directly to human oesophageal tissues. Significantly higher numbers of identifications were made compared to 2D-gel or MALDI-based approaches with selective target identification. The technical reproducibility of TMT labeling and isobaric quantitation was also high. This suggests further biological rather than technical replicates would be required to address the main source of variability.

Reassuringly, at least a fifth of the upregulated proteins in this experiment have prior published evidence of upregulation in OAC. Many of the remainder have an established role in other cancers. Selected down-regulated proteins were also found to have previous published evidence of expression in oesophageal squamous epithelium. This supports the assertion that *bona fide* proteins have been identified as dysregulated using this quantitative proteomic method. The role of novel dysregulated proteins remains to be determined by further work, but the potential for therapeutic approaches is clear.

Candidates were selected for validation by IHC and antibody specificity confirmed. Although only OAC and normal squamous tissue were sampled for the proteomic experiment, protein expression was examined across a range of tissues from the resection specimens.

AGR2 has previously been reported to be expressed in human stomach and other columnar-epithelial tissues with overexpression in Barrett's epithelium and OAC with no expression in squamous epithelia[613-616]. Despite this widespread expression, AGR2 may still be a therapeutic target as it has been reported to regulate cancer cell invasion, potentially via cathepsin-B and D expression[575,617]. Supporting this, AGR2 expression was observed in the metastatic OAC cells in the lymph nodes of patient 15 (Figure 4-8).

The nuclear expression of SAMHD1 and expression in non-epithelial cells in both involved and uninvolved lymph nodes may reflect the reported role of this protein in modulating the immune response in T cells[618]. No role has yet been described for SAMHD1 in cancer and this would be worthy of further study.

GRP-78 has also been previously demonstrated to be upregulated in OAC and plays a key role in the regulation of the unfolded-protein response[594,619]. Using an elegant membrane protein enrichment strategy, a change in localization from the ER to the cell membrane has been noted in cancer cells for several chaperones including GRP-78[374]. This raises the possibility of immunotherapeutic strategies directed against this protein. The staining for GRP-78 appeared to be predominantly cytoplasmic in the sections examined, however, membrane staining is often difficult to determine by IHC especially when cytoplasmic staining co-exists. Further examination of the subcellular expression of GRP-78 in OAC compared to Barrett's, gastric and squamous cells by confocal immunofluorescence microscopy would help to clarify this. A further possibility would be to target GRP-78 directly and specific inhibitors have been described[620].

D4GDI was over-expressed in OAC sections compared to slides from normal squamous tissue. The D4GDI positive cells appeared, however, to be lymphocytes rather than epithelial-derived tumour cells. This expression pattern has been observed for D4GDI with a different antibody in a database of protein expression; the human protein atlas[616] and an early multi-tissue study suggested expression was restricted to haematopoietic cells[621].

Divergent roles for D4GDI have subsequently been proposed in the literature with some evidence for a role in the suppression of metastasis in bladder cancer[622] and in contrast a pro-invasive role in gastric cancer[623]. One previous proteomic study suggested D4-GDI was overexpressed in OAC compared to Barrett's epithelium at both the mRNA and protein level. In contrast to the data presented in this work (Figure 4-9), the IHC images in the published study showed apparently cytoplasmic and membrane staining in epithelial cells with minimal stromal positivity in OAC. The D4GDI antibody used was a "rabbit anti-ARHGDIB antibody at 1:50 dilution

(Abnova Corp.)". Confusingly the paper has been cited in association with a mouse-polyclonal antibody (H00000397-A01) on the Abnova website[624]. No details of the optimization steps taken with this antibody were provided in the publication and it is difficult therefore to confirm the specificity or the identity of the Abnova antibody used.

Variable expression for D4-GDI was noted in the panel of oesophageal cell lines (Figure 4-6) and this may reflect context-dependent regulation. Staining for D4-GDI in a larger number of oesophageal tumours would be required to establish greater confidence over the cell specific expression profile.

The heterogeneous staining pattern for HSP-27 suggests this protein may exhibit dynamic regulation throughout the process of carcinogenesis. A study on breast cancer stem cells identified an association between the stem-cell phenotype and HSP-27 expression. Suppression of HSP27 expression has also been reported to reduce the expression of the EMT-associated markers snail and vimentin[625]. In support of a role for HSP-27 in EMT, stimulation of human peritoneal mesothelial cells to undergo EMT with TGF-β1 increased HSP-27 expression[626]. The same finding was also demonstrated in an independent study on lung cancer cells[627]. Further recent work on prostate cancer cells expanded on this by establishing a requirement for HSP-27 in IL-6 mediated EMT. HSP-27 was also more highly expressed in poorly differentiated prostate cancers, and inhibition of HSP-27 using an anti-sense oligonucleotide reduced the metastasis of prostate cancer cells in a nude mouse xenograft model. Importantly, the same anti-sense oligonucleotide reduced circulating tumour cell counts in a phase I clinical trial in patients with metastatic castration-resistant prostate cancer[628].

HSP-27 has previously been shown to be over-expressed in squamous oesophageal epithelium compared to Barrett's and OAC but no heterogeneity in tumour expression was described in this study[629]. In contrast, HSP-27 mRNA expression increased from non-dysplastic Barrett's samples through dysplasia to adenocarcinoma with the highest levels observed in poorly differentiated

adenocarcinomas[630]. Similarly in OSCC, HSP-27 expression was increased in poorly differentiated relative to moderate or well differentiated tumours[631].

A study on OAC using reverse-phase protein arrays has demonstrated low expression of three isoforms of phosphorylated HSP-27 are associated with a poor prognosis[632]. Interestingly, expression of total HSP-27 appeared to follow the reverse pattern but the findings could not be verified by IHC.

Together these data implicate HSP-27 in tumour progression and metastasis, potentially associated with EMT. This proteomic study provides the first evidence of HSP-27 upregulation in lymph node metastasis in OAC and hints at a potential mechanism of metastasis through dedifferentiation to a more mesenchymal phenotype. Further investigation of the expression of EMT markers and HSP-27 in lymph node metastasis and circulating tumour cells are required in a larger cohort to confirm these findings. If verified, modulation of this pathway could provide a compelling strategy to limit metastasis in OAC.

In conclusion, even with the limited patient numbers in this first study, candidate therapeutic targets have been proposed, confirming the potential power of this comparative proteomic approach. However, the proteome coverage of the pilot study was relatively low when compared to all potential proteins encoded by the human genome. Enhancing proteome coverage may reveal more specific targets and may yield greater insight into the underlying pathways dysregulated in this cancer. Strategies to enable this will be explored in the next chapter.

Chapter 5: Enhancing Oesophageal Tissue Proteome Coverage

5.1 Chapter Aims

a) To evaluate if the altering the tissue lysis method enhances proteome coverage.

b) To determine if a label-free, data-independent proteomic method can enhance proteome coverage compared to data-dependent strategy.

c) To develop a strategy to combine quantitative data from label-free and isobaric labelling (TMT) proteomic experiments.

d) To estimate the probability a protein is not expressed in a sample given incomplete proteomic datasets.

5.2 Introduction

In the previous chapter, quantitative shotgun mass spectrometry was demonstrated to be both accurate and sensitive enough to identify tissue-specific protein expression from human oesophageal biopsies. Although interesting candidate therapeutic targets were proposed, none demonstrated entirely OAC-specific expression. This chapter will explore strategies to improve the number of protein identifications (proteome coverage) from shotgun mass spectrometry experiments. It is predicted that by evaluating greater numbers of candidates there will be a higher probability of identifying tumour-specific proteins.

5.2.1 Intrinsic Limits in the Mass Spectrometry Observable Proteome

The most recent data release of the Human PeptideAtlas encompasses proteins identified from 12,629 unique genes[454]. This global collaborative project has used over 43 million peptide-spectrum matches from 470 experiments to make these identifications. The prior release utilized 14 million peptide-spectrum matches to identify proteins from 11,524 genes. The trebling in data volume required to increment the number of protein identifications by less than 10% suggest the maximum number of proteins identifiable by standard shotgun mass spectrometry methods is being approached. Indeed, there are intrinsic biochemical limits to the number of peptides detectable by mass spectrometry. For example, highly hydrophobic proteins, membrane proteins, and highly basic proteins are all unlikely to be detected by shotgun MS methods[454]. This may explain why the MS detectable proteome is not yet equivalent to the protein-coding genome.

The scanning frequency and mass resolution of a mass spectrometer further restrict the number of detectable peptides from a sample. Fractionation is therefore often employed to reduce sample complexity to enhance coverage. Recent significant advances in instrument technology have produced high mass accuracy, high scanning speed instruments that in combination with orthogonal fractionation method are capable of identifying over 10,500 proteins from a single cell line sample[396,423].

5.2.2 Role of Fractionation

In samples with several milligrams of protein, extensive fractionation methods can be applied to enhance the dynamic range of coverage. One study used several orthogonal methods to identify serum proteins over 12 $\log_{10}$ orders of magnitude in native concentration[402]. Fractionation by OFF-GEL electrophoresis was employed to generate 12 peptide fractions from the pooled samples in the previous chapter. To enhance this method, further separation of peptide mixtures could be achieved by increasing the number of OFF-GEL fractions to the current limit of 24.

5.2.3 Use of Detergents for Tissue Lysis

Enhancing proteome solubilisation could further improve proteome coverage. The previous study used a lysis buffer with 0.05% (v/v) SDS; the maximum detergent concentration compatible with direct tryptic digestion. Higher detergent concentrations could enable the solubilisation of more hydrophobic proteins but in addition to creating problems with enzymatic digestion, SDS and many other detergents ionise readily and can corrupt peptide mass spectra. The lysis buffer can, however, be exchanged to a trypsin and MS compatible solution using molecular weight cut-off filters. An application of this technique has been termed Filter-Aided Sample Preparation (FASP)[443].

5.2.4 Optimising the Detection of Tumour-Specific Proteins

The expression in gastric epithelium of some of the proteins over-expressed in OAC compared to normal squamous epithelium highlights a limitation in the original experimental design. Identified proteins with a high TvO ratio may be expressed in all glandular rather than squamous epithelia, may be specifically under-expressed in squamous epithelia, or may be specifically expressed in adenocarcinoma tissue. Tractable therapeutic candidates may be more likely to be identified in the latter category but using the TvO ratio alone, they cannot be discriminated from proteins in the other categories.

If OAC, normal oesophagus and gastric tissue from the same patient were analysed together, protein expression could be defined by both tumour vs. normal oesophageal (TvO) and tumour vs. normal gastric (TvG) comparisons. Using this method,

proteins selectively upregulated in OAC could be more easily identified since those proteins expressed in all columnar epithelial tissues should be expressed at similar levels in both gastric and OAC tissue or at higher levels in gastric tissue.

5.2.5 A Systematic Sequencing Approach

Shotgun proteomic experiments most commonly use a data-dependent workflow. This requires that precursor (peptide) ions be detected in the MS1 phase[386]. For the pilot study, this approach was followed and a maximum of the top three precursor ions in each MS1 scan were selected for fragmentation and detection in the MS2 phase. There are limitations in the data-dependent method as all precursor ions cannot be selected for fragmentation in the time scale of a standard chromatographic peak. Similarly, some peptides ions are below the limit of detection and are never selected for fragmentation[405].

A possible method to overcome these limitations is to define precursor ion m/z ranges (windows) and to fragment and sequence all the precursor ions within the m/z range. By systematically screening through large precursor m/z ranges, peptides from a potentially wider dynamic range of expression can be identified. This data-independent approach has been termed Precursor Acquisition Independent From Ion Count (PAcIFIC)[404]. A limitation of the method is the requirement to perform multiple LC runs to cover the full m/z range. Improvements in instrument sequencing speeds and optimized LC gradients have reduced the protocol time from over 4 days to just over 24 hours per sample[406].

5.3 Results

5.3.1 Evaluation of Proteome Extraction Methods

As the FASP method has not previously been applied to oesophageal tissue, the tissue lysis protocol was first optimized using snap frozen mouse small intestinal tissue to provide homogeneous, reproducible tissue samples with a similar composition to human gastrointestinal tissue. Buffer compositions and the steps from the FASP protocol were applied as published with minor modifications to suit local equipment. Differing sequences of lysis buffer addition and homogenization

were initially tested. From this initial experiment, a biopsy weight of at least 2 mg and a ratio of at least 15:1 lysis buffer (μL) : biopsy weight (mg) appeared to yield the most protein (Table 5-1). The addition of needle sonication for 30s after biopsy lysis increased the final protein yield after buffer exchange by approximately two fold (data not shown).

The published FASP protocol describes on-column tryptic digestion followed by centrifugation to collect the resultant tryptic peptides. To evaluate this method, two mouse small intestine biopsies (A, 28.9 mg and B, 30.2 mg) were homogenised dry and dissolved in complete lysis buffer added at a 20:1 ratio. Lysates were then sonicated and 30 μL of clarified lysate buffer-exchanged, alkylated and diluted in TEAB as per the FASP protocol.

Lysates were digested on-column by overnight incubation with trypsin and then centrifuged. The protein concentrations of the NMWCO filter residuate and filtrate revealed the majority of the protein mass was retained in the filter (Table 5-2). The method was therefore modified so that the residuate was aspirated and digested off-column, resulting in an improved yield (Table 5-2). Off-column digestion was subsequently used for all human tissue proteomic experiments.

The original FASP protocol specified the use of 10 kDa NMWCO filters. A subsequent investigation revealed minimal protein loss but a significant reduction in protocol time with the use of 30 kDa filters[633]. To evaluate if significant protein loss occurs during FASP and to confirm the efficacy of tryptic digestion, the off-column tryptic digests detailed in Table 5-2 and the corresponding buffer exchanged lysates were resolved by SDS-PAGE. Images of the Coomassie-stained gels confirm significant concentrations of proteins <30 kDa are retained during the FASP procedure and the off-column trypsin digestion generates an almost complete digestion (Figure 5-1). The enhanced yield from sample sonication is also evident.

Table 5-1 Optimisation of Tissue Lysis Protocol. Protein concentration was determined using the RC-DC assay after lysate sonication and buffer exchange.

Mouse Small Intestine Biopsy (Bx)	Biopsy weight (mg)	Volume of lysis buffer added pre homogenisation (µL)	Volume of 10% (v/v) SDS or buffer added post homogenisation (µL)	Final buffer volume (µL)	Final Tissue : Buffer ratio	Protein Concentration (µg/µL)	Total Protein Yield (µg)	**Yield (protein (µg) /biopsy mass (mg))**
Small Bx 1:5	24.9	125 DFLB	83	208	1:7.5	Minimal	x	**X**
Medium Bx 1:5	35.5	178 DFLB	118	296	1:7.5	1.13	333	**9.4**
Large Bx 1:5	38.6	193 DFLB	127	320	1:7.5	1.14	365	**9.4**
Small Bx 1:10	26.5	265 DFLB	176	441	1:15	0.17	75	**2.8**
Medium Bx 1:10	31.7	317 DFLB	209	526	1:15	2.09	1099	**34.7**
Large Bx 1:10	41.8	418 DFLB	275	693	1:15	1.23	852	**20.4**
Small Bx 1:20	11.7	234 DFLB	154	388	1:30	0.38	147	**12.6**
Medium Bx 1:20	26.4	528 DFLB	348	876	1:30	1.06	929	**35.2**
Large Bx 1:20	34.4	688 DFLB	454	1142	1:30	0.77	879	**25.5**
Dry + Small Bx	21.1	X	211 SDSB	211	1:10	1.68	354	**16.8**
Dry + Medium Bx	35.9	X	359 SDSB	359	1:10	2.07	743	**20.7**
Dry + Large Bx	61.8	X	618 SDSB	618	1:10	2.11	1303	**21.1**
Medium Bx 1:10	27.4	274 SDSB	X	274	1:10	1.49	408	**14.9**
SDS buffer only	x	200 SDSB	X	200	1:10	0.0	x	**X**

DFLB – Detergent-free Lysis buffer, SDSB – Complete lysis buffer (DFLB + 4% SDS)

Table 5-2 Comparison of Digestion Conditions for two Mouse Small Intestine Biopsies

Starting Tissue Sample	Sonicated	Tryptic Digestion Condition	Protein Concentration (mg/ml)
A	Y	Filtrate	**0.062**
A	Y	Residuate	**0.144**
A	Y	Off-Column	**0.504**
B	Y	Filtrate	**0.058**
B	Y	Residuate	**0.109**
B	Y	Off-Column	**0.471**

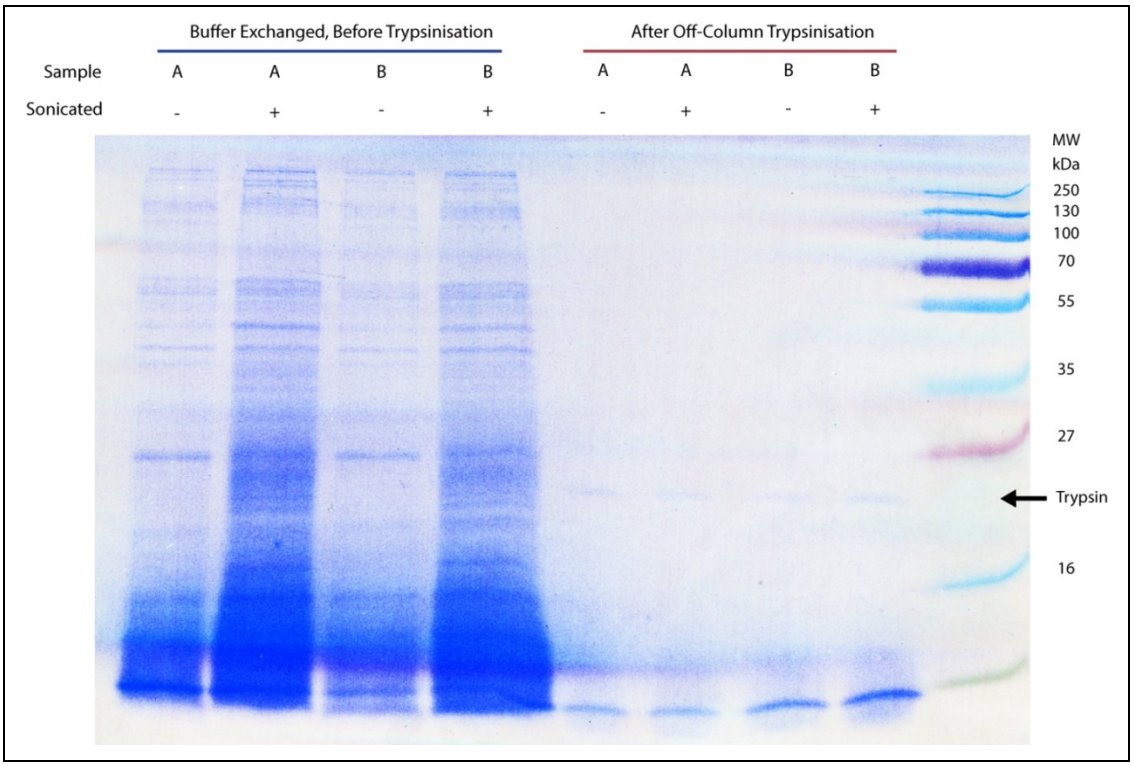

Figure 5-1 Effect of Buffer-Exchange, Sonication and Tryptic Digestion on Protein Size Distribution. Lysates were buffer exchanged, alkylated, and diluted using the modified FASP protocol. Samples before and after off-column tryptic digestion were resolved by SDS-PAGE (16µL of undiluted lysate and 4 µL of 5x Sample Loading Buffer per lane, 10% agarose), and the gel stained with Coomassie blue before scanning. The highlighted band in the Tryptic digests is presumably trypsin (predicted molecular weight (MW) 24 kDa).

5.3.2 Comparison of modified FASP (mFASP) and initial methods.

To evaluate the effect of this change in lysis method on proteome coverage, the modified FASP protocol was applied to samples from two further patients (44 and 46). To improve the probability of identifying tissue-specific proteins, patient matched OAC, normal oesophageal and gastric tissue were all lysed and processed using the modified FASP (mFASP) protocol before TMT-labeling. Peptides from each tissue type were labeled in technical duplicate, incorporating all six TMT reporters and therefore analyzing all the tissues from one patient together in the same LC and mass spectrometry run. Technical replicate labeling was incorporated to allow the degree of technical variation to be determined on an individual sample basis. To attempt to further enhance proteome coverage, labeled peptides from all the tissues from each patient were pooled and then 24 fractions resolved by OFF-GEL electrophoresis. As this increased the time to run a single patient's samples to over 24 hours, only a single LC-MS/MS run was performed for each fraction and the data for all fractions pooled. Protein identifications were made as previously using the Easyprot platform. The initial and modified TMT workflows are shown in Figure 5-2.

The total number of protein IDs and their contributions from the initial experiment and the subsequent mFASP TMT experiments are demonstrated in Figure 5-3. This shows that only 110 of the total 2168 proteins were unique to the initial experiment and that 1,542 proteins were detected in at least two experiments. The combination of the mFASP method and the increased number of OFFGEL fractions appears, therefore, to significantly increase the number of protein identifications.

5.3.3 Comparison of PAcIFIC and TMT shotgun proteomic methods

To evaluate if a systematic shotgun proteomic sequencing method could further enhance proteome coverage, PAcIFIC LC-MS/MS was applied to tryptic peptides generated using the mFASP method. Eleven days of mass spectrometry time were available for this analysis. Although it was anticipated that the additional analysis of patient-matched gastric tissue along with OAC and normal oesophageal tissue would

aid the identification of tumour specific proteins, this would limit the PAcIFIC analysis to 3 patients. Biological variability was shown to be a potentially significant factor in the initial TMT experiment and therefore OAC and normal oesophageal tissue from five patients were analysed. The PAcIFIC workflow is summarised in Figure 5-4.

A total of 3,019 proteins were identified across all PAcIFIC samples. The PAcIFIC mass spectra were searched against an International Protein Index (IPI) database and therefore proteins were denoted with an IPI reference number. The TMT data, however, were searched against the Swissprot curated database and proteins were denoted with a Uniprot Accession number (AC). Although different databases were searched, both TMT and PAcIFIC experiments used conservative identification strategies requiring two different peptides per protein and maintaining an FDR of less than 1%. To compare experiments, the PAcIFIC IPI reference numbers were mapped to 2,936 unique AC identifiers. In comparison, the mFASP TMT experiments yielded 2,058 unique AC identifications with 1,483 proteins identified in both experiments.

The individual TMT and PAcIFIC experiments are compared in Table 5-3. The highest numbers of protein identifications were observed in the mFASP TMT experiments. TMT quantification appears to offer a significantly greater quantitative dynamic range than label-free quantitation using spectral counting.

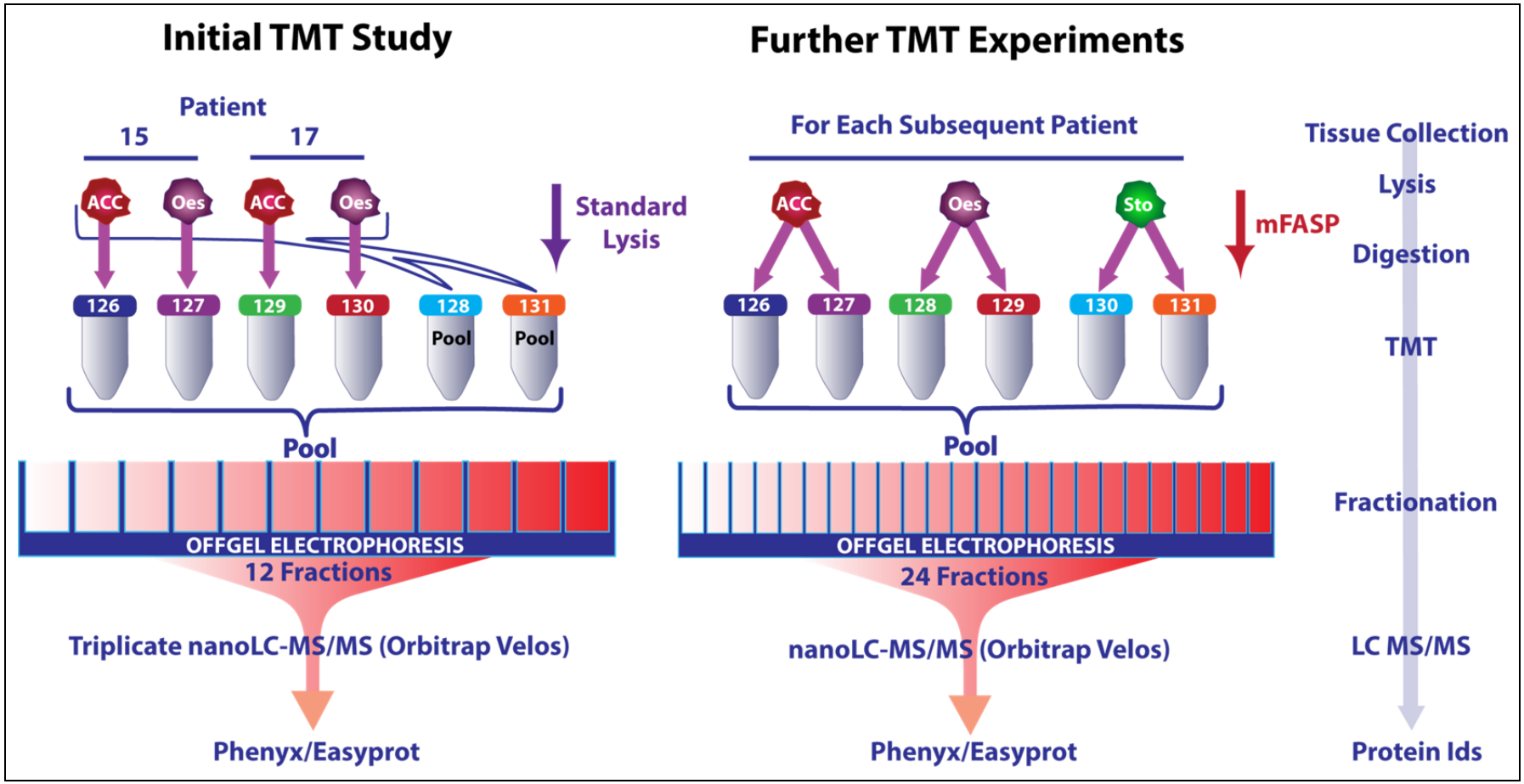

Figure 5-2 Summary Experimental Protocol for TMT Proteomic Experiments. Tryptic peptides from each tissue type were labelled with TMT reporters as demonstrated. The protocol on the left was followed for the pilot TMT experiment and the right for all subsequent experiments. Abbreviations: ACC – Adenocarcinoma biopsies, Oes – normal squamous oesophageal biopsies, Sto – normal gastric biopsies, mFASP – modified Filter-Aided Sample Preparation. TMT – Tandem Mass Tags. LC – Liquid Chromatography, MS/MS – Tandem Mass Spectrometry.

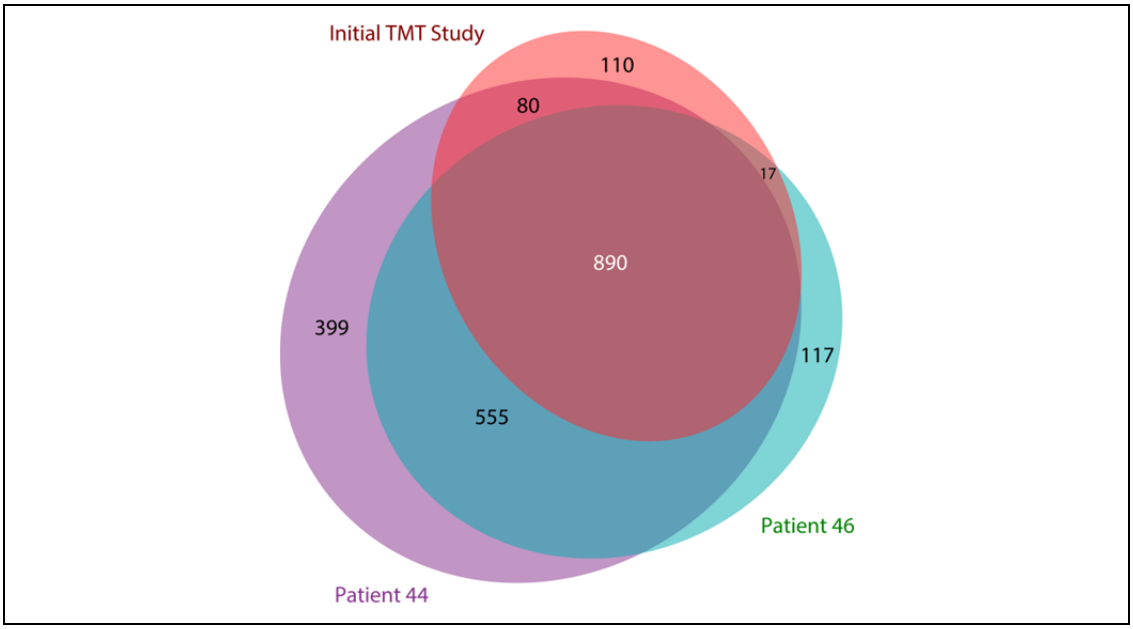

Figure 5-3 Overlap of Protein Identifications between Initial (Red) and mFASP TMT experiments (purple and green).

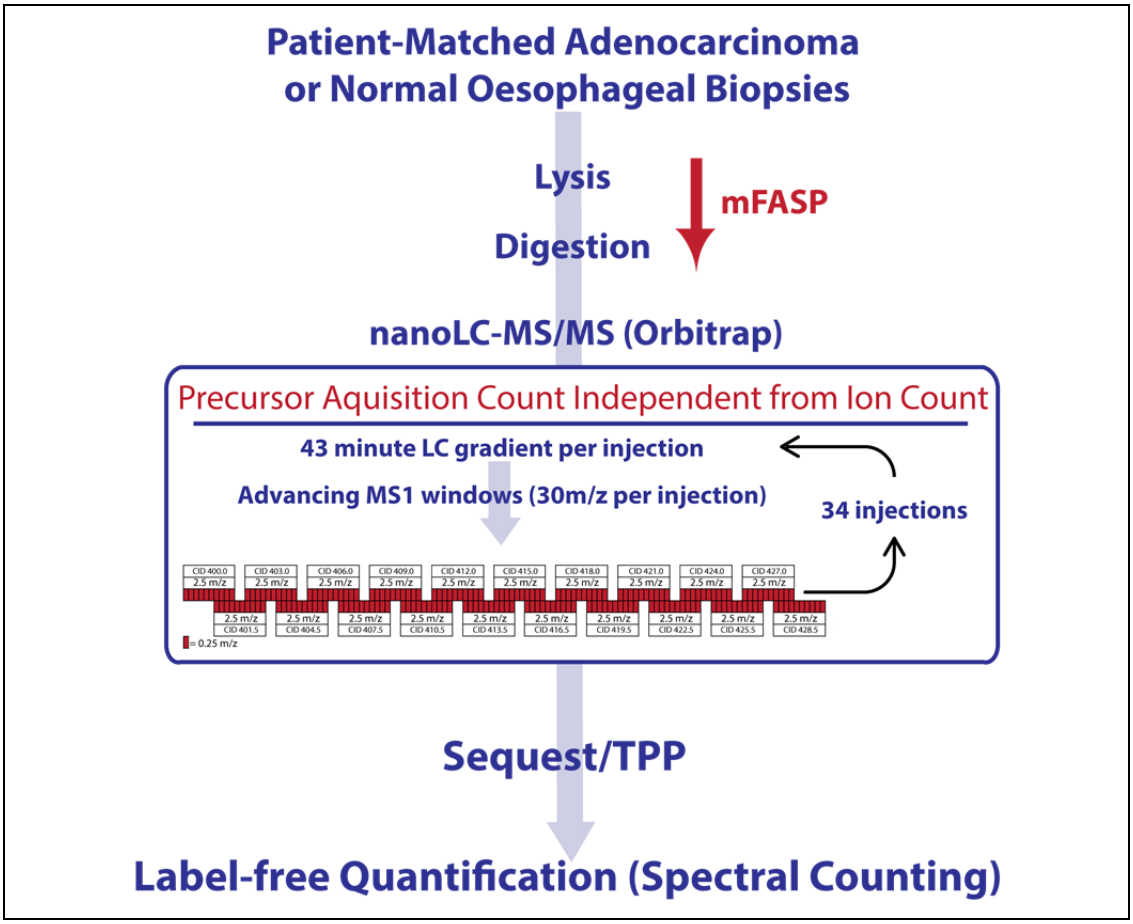

Figure 5-4 Summary of PAcIFIC Workflow. Samples were processed using the mFASP method and then analysed using by PAcIFIC LC-MS/MS. Proteins were identified and quantified using the Trans-Proteomic Pipeline (TPP) suite of tools.

Table 5-3 Comparison of Protein Identifications and Quantitative Dynamic Range of Different Proteomic Methods

Patient	Tissue Samples	Lysis Method	MS Method	Fractionation Method	Number of Proteins (FDR <1%, 2 unique peptides)	Quantitative Dynamic Range (Log_{10})
15+17	T,N	Standard	TMT	OFFGEL, 12 fractions	1097	6.7
44	T,N,G	mFASP	TMT	OFFGEL, 24 fractions	1924	6.9
46	T,N,G	mFASP	TMT	OFFGEL, 24 fractions	1579	7.0
23	T	mFASP	PAcIFIC	Nil	1879	2.25
23	N	mFASP	PAcIFIC	Nil	1183	2.48
26	T	mFASP	PAcIFIC	Nil	1408	2.21
26	N	mFASP	PAcIFIC	Nil	923	2.34
27	T	mFASP	PAcIFIC	Nil	1585	2.23
27	N	mFASP	PAcIFIC	Nil	1095	2.43
32	T	mFASP	PAcIFIC	Nil	1471	2.29
32	N	mFASP	PAcIFIC	Nil	1050	2.49
45	T	mFASP	PAcIFIC	Nil	1003	2.37
45	N	mFASP	PAcIFIC	Nil	977	2.31

Quantitative Dynamic Range was calculated as the highest TMT reporter ion intensity divided by the lowest or the highest spectral count divided by the lowest spectral count. T – Tumour, N- Normal Oesophagus, G- Normal Gastric.

5.3.4 The Effect of the mFASP protocol on protein solubilisation.

To explore if the mFASP method improved proteome coverage by enhancing membrane protein solubilisation, the primary subcellular loci for proteins identified via each experimental method were compared. Summary localization data were available for 921 out of 1,097 proteins (84%) identified in the pilot TMT study and 1,732 out of 2,058 proteins (84%) in the mFASP TMT data. For the PAcIFIC experiments, 1,963 of the 2,936 proteins mapped to Uniprot Accessions had localisation data (67%). The primary subcellular localisation for proteins identified following each experimental method are summarised in Figure 5-5.

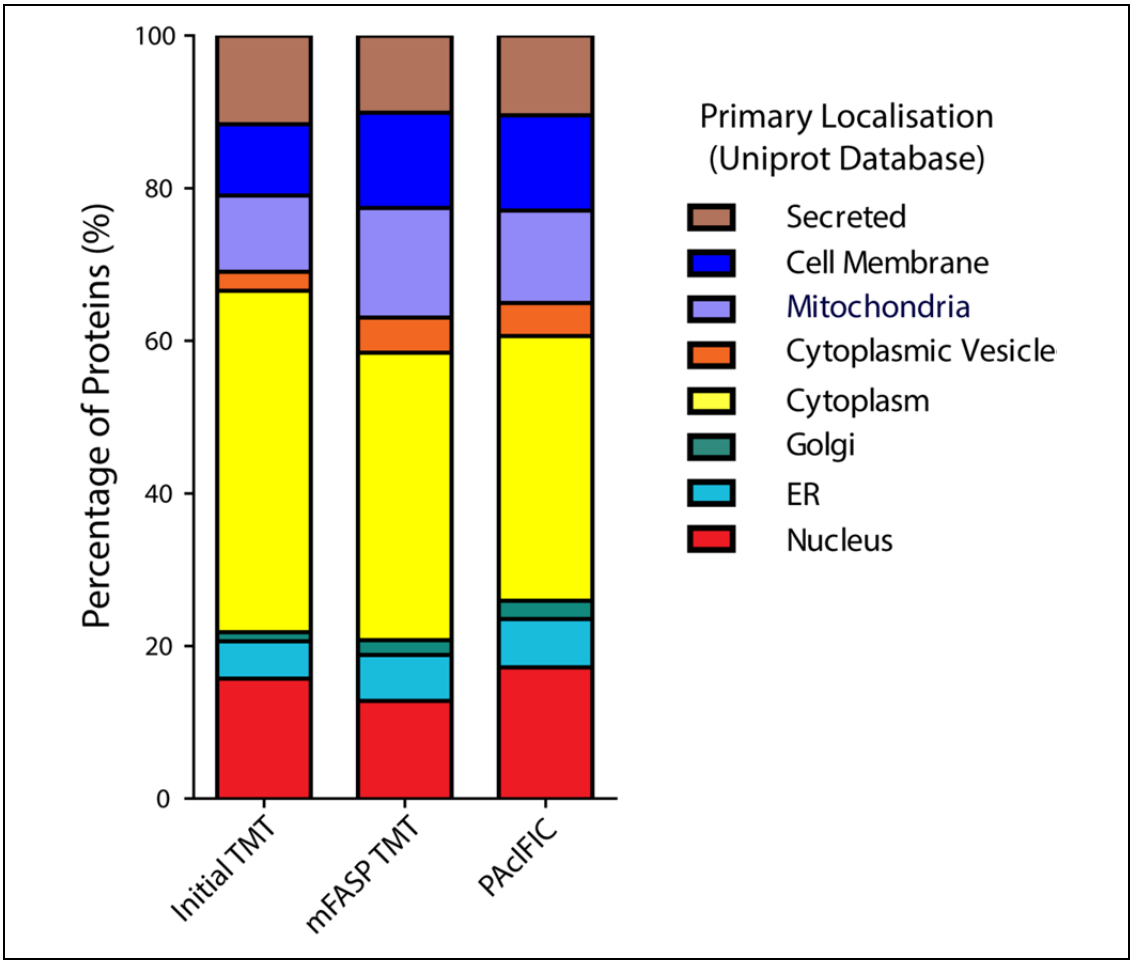

Figure 5-5 Summary of Primary Subcellular Localisation of Proteins Identified via Each Proteomic Method. See Methods 2.11.5 for mapping details.

This demonstrated that the mFASP method produced an increase in the identification of cell membrane proteins and proteins primarily localized to subcellular membrane bound structures; cytoplasmic vesicles (lysosomes, endosomes, peroxisomes, etc.), the ER, golgi and mitochondria. Based on this limited analysis, this suggests that the mFASP method does indeed enhance the solubilisation of membrane proteins.

5.3.5 Combining Proteomic Datasets

Seven patients' samples had been processed using the mFASP method. Clinical data for these patients are summarized in Table 5-4. Combining the results from mFASP TMT and PAcIFIC experiments may enhance the probability of identifying tumour specific proteins.

Table 5-4 Clinical Characteristics of Patients Donating Samples for mFASP Proteomic Experiments.

	Patient						
	23	26	27	32	45	44	46
Experiment	PAcIFIC	PAcIFIC	PAcIFIC	PAcIFIC	PAcIFIC	mFASP TMT	mFASP TMT
Gender	Male	Male	Male	Male	Male	Male	Male
Age	57	53	60	57	78	59	67
Histology	ACC	ACC	ACC	ACC	ACC	ACC	ACC
Location	OGJ Type I	OGJ Type I	Lower Oes	OGJ Type I	OGJ Type I	OGJ Type II	OGJ Type I
Neoadjuvant Therapy	2xCF	2xCF	2xCF	2xCF	Nil	2xCF	2xCF
Post-Chemotherapy CT	Stable Disease	Stable Disease	Stable Disease	Stable Disease	Stable Disease	Stable Disease	Stable Disease
Surgery	ILO	ILO	ILO	ILO	ILO	ILO	ILO
Maximum Tumour Diameter	46mm	35mm	40mm	94mm	40mm	38mm	70mm
Distance to PRM and DRM	>1mm	>1mm	>1mm	>1mm	>1mm	>1mm	>1mm
Distance to CRM	5.5mm	0.9mm	1.0mm	0.1mm	0.0mm	4.2mm	0.0mm
Differentiation	Moderate	Poor	Poor	Poor	Poor	Moderate	Poor
Lymphovascular Invasion	Y	N	Y	Y	Y	Y	Y
Venous Invasion	Y	N	N	Y	Y	N	N
Perineural Invasion	N	Y	N	Y	Y	N	Y
T stage	T2	T3	T3	T3	T3	T2	T4a
Number of Positive Nodes	0	2	7	7	6	2	8
Total Number Nodes Resected	43	14	20	19	16	27	18
AJCC Stage	IB	IIIA	IIIC	IIIC	IIIB	IIB	IIIC
Alive at Time of Analysis	Yes	Yes	Yes	Yes	No	Yes	No
Overall Survival From Diagnosis	13 months (censored)	25 months (censored)	24 months (censored)	24 months (censored)	9 months	18 months (censored)	15.3 months

Abbreviations. , 2xCF - 2 cycles of Cisplatin and 5-Fluorouracil, ILO – Ivor-Lewis Oesophagectomy, AJCC – American Joint Committee on Cancer, OGJ – Oesophagogastric Junctional Tumour, PRM – Proximal resection margin, DRM – Distal resection margin, CRM – Circumferential resection margin, Y-Yes, N-No.

5.3.6 Generating a Matrix of Unique Quantified Proteins

The protein identifications (IDs) from TMT and PAcIFIC experiments were first combined into a list of unique Uniprot ACs with corresponding abundance data. Proteins from 3,549 unique genes were identified and quantified in OAC or normal oesophageal samples from at least one of the seven patients. It is apparent from Table 5-3 that each proteomic experiment only measured a subset of the proteins in the tissue lysate. In the total set of identified proteins each protein may therefore

have been identified (and quantified) between one and seven times. The differing scales, dynamic range and quantitation methods preclude combining values for proteins across all experiments by arithmetic methods.

5.3.7 Combining quantitative proteomic datasets with overlapping protein identifications

Strategies for combining quantitative expression datasets have been developed for application in microarray analysis. A popular method is quantile normalization, where proteins are ranked within an experiment based on their abundance and this rank replaced with the corresponding value calculated from the mean of the rank-ordered samples[634]. This method normalises the experimental distributions and allows experiments with differing quantitation methods to be combined, enhancing the power of the analysis. Proteins not observed in an experiment cannot, however, be ranked by quantile normalization.

The ratio of protein expression between tissues is scale-independent. This ratio could be used to compare relative protein expression between tumour and normal samples across PAcIFIC and TMT experiments. To generate this combined expression value, only tumour vs. normal oesophageal expression ratios can be used, since gastric tissues were not assessed by PAcIFIC. Ratios, however, also require quantitation values for both samples. When the peptide reporter ratios were combined for each protein for the TMT experiments, no tissue specific expression was observed and so ratios were calculable for all proteins. Conversely, in the PAcIFIC experiments, over a quarter of the identified proteins were only observed in the OAC samples. Dealing with missing observations is a central problem in proteomic informatics and will be considered next.

5.3.8 Dealing with Missing Observations

There are two situations where missing values occur in the combined proteomic dataset;

(1) A protein is not observed in any experiment for a tissue type. "7 zero" = 7z

(2) A protein is observed in some of the PAcIFIC experiments. "False negative" = FNeg

To consider these situations some assumptions will be made:

(a) Each sample from a tissue type is effectively a technical replicate (biological variation is ignored). More specifically each tissue sample is regarded to contain very similar proteins and there are very few experiment-specific proteins.

(b) A subset of the total human (normal) proteome will be expressed in each tissue type (c.f. all tissues expressing all proteins.). A tissue type may therefore contain tissue-specific proteins.

(c) There is a biochemical bias in peptide and therefore protein identification by mass spectrometry. This bias is assumed to be equivalent in all tissues so that there is an equal probability of a protein being identified from one tissue compared to another. When a protein is not detected in any of the normal tissue experiments but detected in many or all of the tumour tissue experiments the protein is more likely to be truly not expressed in normal tissue (a true negative) than if the protein is only detected in one of the tumour samples.

5.3.9 The False Negative (FNeg)

In situation 2 and following assumption a, it can be concluded that if a protein has been observed once in a tissue sample, all subsequent missed identifications must be false negatives. The true expression value could be estimated for that tissue by using the observed values.

The relationship between the probability of a false negative and protein abundance was assessed. The false negative rate was calculated on a per protein basis for tumour and normal tissue. This was expressed as;

$$Terr = 1 - Ft$$

$$Nerr = 1 - Fn$$

Where;

Terr = the false negative rate in tumour

Nerr = the false negative rate in normal oesophagus

Ft = the frequency of observation of a protein in the seven experiments for tumour tissues.

Fn= the frequency of observation in normal tissues.

Terr and Nerr can take one of 8 discrete values from 0 – 1 (0/7, 1/7, 2/7 … 6/7, 7/7). The case of Terr or Nerr = 1 will be considered subsequently as the 7 zero problem.

To assess the relationship between Nerr, Terr and protein abundance, data from all seven experiments were evaluated. Protein abundance was estimated from the geometric mean of peptide reporter intensities from the TMT experiments as the dynamic range and accuracy were considered superior to spectral counts (Table 5-3).

Surprisingly, when protein abundance was plotted against Ft (1-Terr) and Fn (1-Nerr), there appeared to be no relationship (Figure 5-6A). This suggests that protein identification was not biased by abundance. The probability of each identification rate was also calculated for each tissue type (Figure 5-6B). A bimodal distribution for identification frequency was identified for each tissue type, with a group of ubiquitously identified proteins in both tumour and normal tissues. There also appeared to be a bias towards OAC protein identification with fewer proteins identified exclusively in normal oesophagus. This may reflect the higher numbers of protein identifications in the PAcIFIC tumour samples compared to the normal

oesophageal samples (Table 5-3). To examine the relationship between Ft and Fn the conditional probability of Ft for a given Fn and vice versa were plotted (Figure 5-7).

There was a good correlation between Ft and Fn providing some support for the assumption (c) that tumour and normal proteins behave in a similar fashion in MS experiments.

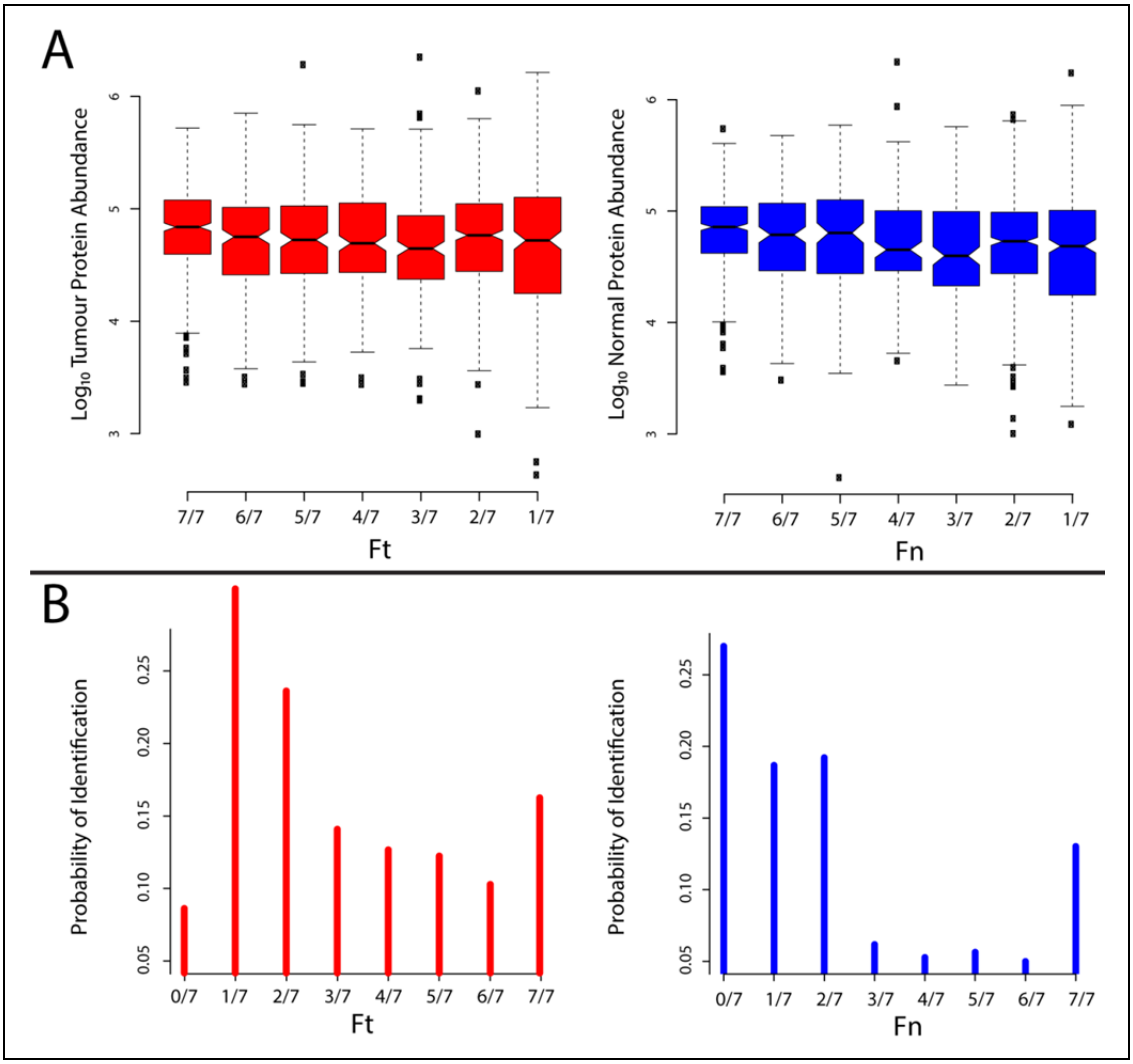

Figure 5-6 (A) Notched boxplots demonstrating protein abundance by frequency of identification in tumour (Ft) and normal oesophageal (Fn) tissues. The whiskers represent the maximum or minimum values or 1.5 times the inter-quartile range if outliers (circles) extend beyond this. The notches approximate 95% confidence intervals of the median[635]. (B) Relative proportion of proteins identified at each frequency.

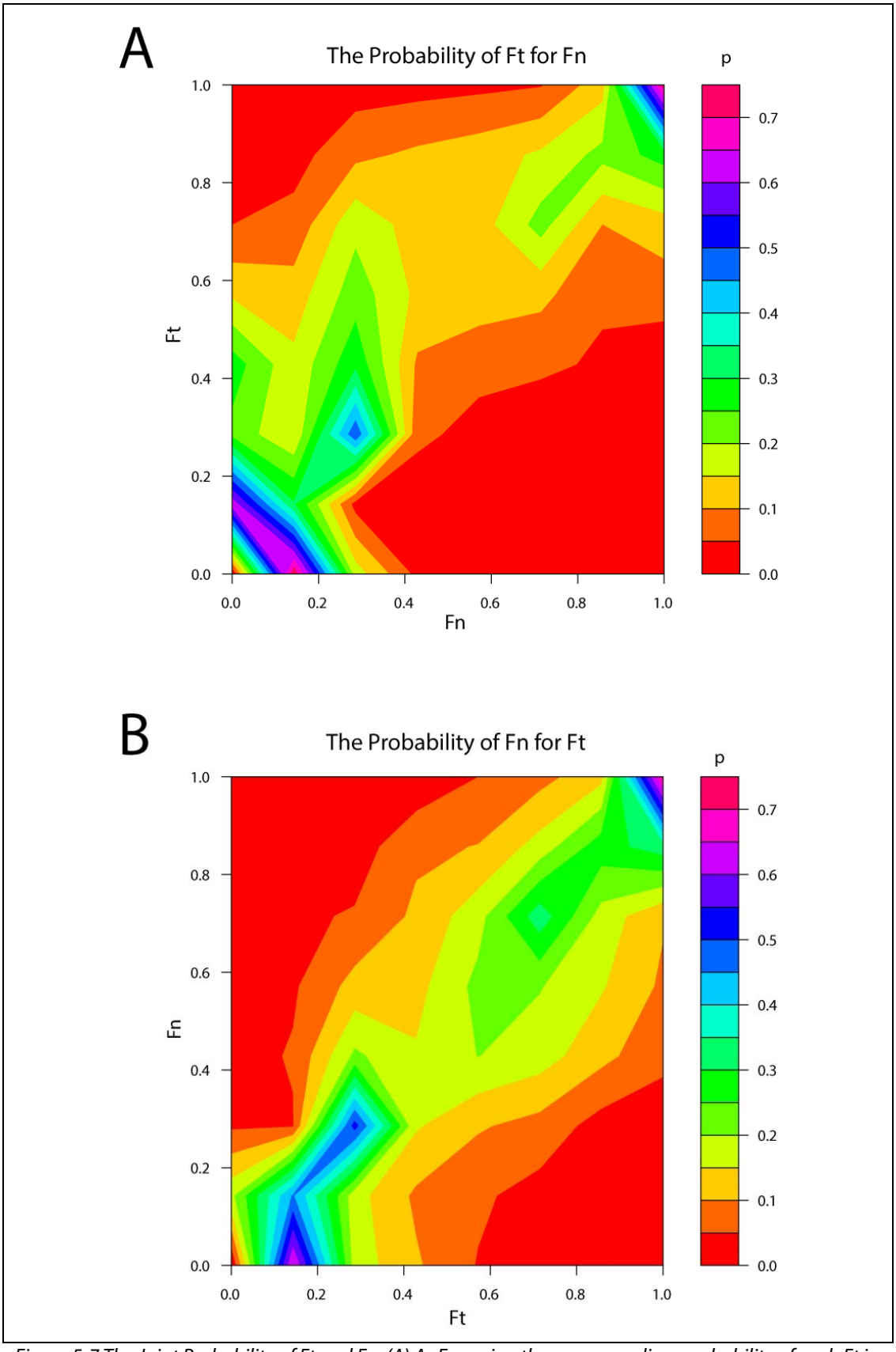

Figure 5-7 The Joint Probability of Ft and Fn. (A) As Fn varies, the corresponding probability of each Ft is displayed using the colour scale. (B) As Ft varies the corresponding probability of each Fn is displayed.

5.3.10 A Statistical Framework to Identify True Negatives

In situation (1) and from assumption (b), not observing a protein in seven experiments (7z) may represent a true or false negative. From assumption (c) if Fn=0 then as Ft approaches 1 the probability of a true negative is more likely. An estimate of the probability of a true negative given the Ft (or Fn) was therefore sought as this may allow false negative data to be excluded.

Considering several conditions, initially in normal tissue;

NE = a protein is not expressed in the normal tissue. Observing 7z for this protein would be a true negative.

Fn = the frequency of identification in the normal tissue

Fx = the frequency of identification across tissues – ie. a co-ordinate Ft,Fn

The probability that a protein is not expressed in normal tissue given an Ft, Fn co-ordinate can be expressed as $p(NE|Fx)$. This is the true negative probability and will allow an Ft threshold to be determined above which proteins will be considered to not be expressed in the normal tissue i.e. tumour specific. The data for proteins with Fn = 0 and an Ft below this threshold will be discarded.

Expanding this expression using Bayes theorem[636] gives;

$$p(NE|Fx) = p(F_x|NE) * \left(\frac{pNE}{pFx}\right)$$

Equation 10

Defining each component of this equation;

$p(F_x|NE)$ = the probability of obtaining the Ft,Fn co-ordinate given that the protein is not expressed in normal tissue.

The probability of a protein not being expressed is zero when the protein has been observed at least once. For all values of Ft,Fn where Fn ≠ 0, $p(F_x|NE)$ will be 0 and, from Equation 10, $p(NE|Fx)$ will also be zero.

pFx = the probability of obtaining the given Ft, Fn co-ordinate across the tumour and normal oesophageal proteomes (when normal protein databases are used). The size of these proteomes are unknown but a conservative strategy would be to take the best estimate of the current total MS observable proteome (Y) from the human PeptideAtlas; 12,629 proteins[423]. Therefore pFx can be estimated as the total number of proteins with Ft,Fn divided by 12,629.

pNE = the probability that a randomly selected protein is not expressed in normal tissue. This can be defined as 1-(N/Y) where N is the total number of proteins expressed in the normal oesophageal tissue.

As true negatives can only be identified if values of pFx are $Ft_{>0/7}$,$Fn_{0/7}$ or $Fn_{>0/7}$,$Ft_{0/7}$ only these proteins are considered. The probability of a specific Ft given $Fn_{0/7}$ in our data set is demonstrated in Table 5-5.

Table 5-5 The probability of Ft given $Fn_{0/7}$

Ft	$pFt\|Fn_{0/7}$
0	0
1/7	0.6534
2/7	0.1775
3/7	0.1065
4/7	0.0470
5/7	0.0157
6/7	0
7/7	0

The probability of the co-ordinate combination $Ft,Fn_{0/7}$ can be considered as the sum of the probability of achieving the combination when the protein is not expressed and the probability of achieving the combination when the protein is expressed but has been missed.

(ie. Total probability = true negative probability + false negative probability)

This is expressed algebraically for the combination $Ft,Fn_{0/7}$ as;

$$p\left(Ft, Fn_{0/7}\right) = \left[p\left(Ft, Fn_{0/7}\middle|NE\right) \times pNE\right] + \left[p\left(Ft, Fn_{0/7}\middle|Mn\right) \times pMn\right]$$

Equation 11

Where;

$p\left(Ft, Fn_{0/7}\middle|Mn\right)$ = the probability of obtaining an Ft, $Fn_{0/7}$ combination given that the protein has been missed (not detected) in normal tissue.

pMn = the probability that a protein has been missed in the normal tissue.

5.3.11 The Effect of Relative Proteome Size on pFt and pFn

pNE is proportional to the relative sizes of the tumour proteome and normal proteome.

This is illustrated in Figure 5-8;

$$pNE = \left(\frac{T}{T+N}\right) \qquad pMn = \left(\frac{N}{T+N}\right)$$

Intuitively proteins with an $Ft_{1/7}Fn_{0/7}$ and an $Ft_{2/7}Fn_{0/7}$ are expected to be predominantly in the "N" area and proteins with an $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$ to be predominantly in the "T" area".

In the example in Figure 5-8A, N= 400 proteins and T = 600 proteins. For the purposes of this example, following assumption (c), it is suggested that 50% of proteins in N have an $Ft_{1/7}Fn_{0/7}$ and $Ft_{2/7}Fn_{0/7}$ and 5% have an $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$. In contrast it is proposed that 10% of proteins in T have an $Ft_{1/7}Fn_{0/7}$ and $Ft_{2/7}Fn_{0/7}$ and 50% of T have an $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$.

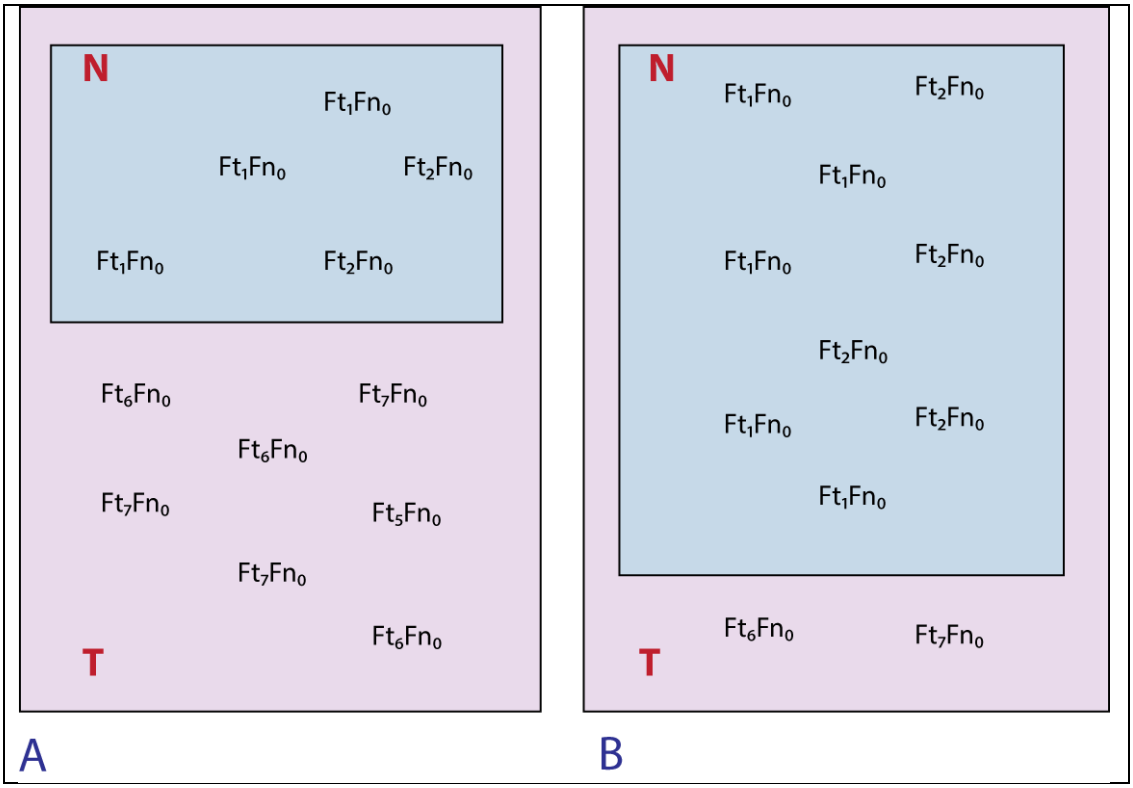

Figure 5-8 Hypothetical Examples of (A) High and (B) Low Mean Ft. The outer area (T) represents proteins only expressed in tumour tissue and the inner area (N) represents proteins expressed in both tumour and normal tissues.

The probability of $Ft_{1/7}Fn_{0/7}$ and $Ft_{2/7}Fn_{0/7}$ for this hypothetical dataset (T+N = 1000 proteins) can be calculated from Equation 11 as;

$(0.1 \times 0.6) + (0.5 \times 0.4)$

$= 0.26$

In the same fashion, the total probability for $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$ is calculated as;

$(0.5 \times 0.6) + (0.05 \times 0.4)$

$= 0.32$

So with an $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$ in this experiment and using the probabilities provided, the true negative probability can be calculated. This is the probability the protein is in T, given the observed data (F_x); $p(NE|Fx)$

From Equation 10;

$$p(NE|Fx) = \frac{p(Ft,Fn_{0/7}|NE) \times pNE}{p(Ft,Fn_{0/7})}$$

Which for Fx = $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$

= 0.3/0.32

= 0.9375

In this example, these proteins would be considered likely true negatives (not expressed in the normal tissue).

In contrast when Fx = $Ft_{1/7}Fn_{0/7}$ or $Ft_{2/7}Fn_{0/7}$;

$p(NE|Fx)$ = 0.06/0.2

= 0.3

In this example there is a reasonable probability the proteins are false negatives (are expressed in normal tissue but have been missed) and the data would be discarded.

In the second illustration (Figure 5-8 B) values of N = 800 proteins, T= 200 proteins are proposed with the same probabilities of $Ft_{1/7}Fn_{0/7}$ or $Ft_{2/7}Fn_{0/7}$ and $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$ for each tissue type.

The total probability of $Ft_{1/7}Fn_{0/7}$ or $Ft_{2/7}Fn_{0/7}$ would be;

= (0.1*0.2) + (0.5*0.8)

= 0.42

and for $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$

= (0.5*0.2) + (0.05*0.8)

= 0.14

Therefore $p(NE|Fx)$ when Fx= $Ft_{6/7}Fn_{0/7}$ or $Ft_{7/7}Fn_{0/7}$;

= 0.714

and when Fx = Ft_1Fn_0 or Ft_2Fn_0 ;

= 0.05

The loss of predictive power in the second example is due to the larger set of proteins with a higher false negative probability (Ft_1Fn_0 of Ft_2Fn_0). In situation (1) if assumption (c) holds true it can be appreciated that the mean $Ft|Fn_{0/7}$ will directly correlate with the proportion of false negative proteins.

This generalisation is supported if a theoretical complete sampling of tumour and normal oesophageal proteomes is imagined. In this scenario the false negative rate should tend towards zero and the average Ft for an Fn=0 will tend towards 1. In the converse situation with a sparse dataset there will be many examples of proteins with an Ft_1Fn_0 or Ft_2Fn_0 and it will be unclear if they are true or false negatives.

5.3.12 Derivation of an Estimate of False Negative Probability

Expanding the false negative probability term $p(Ft, Fn_0|Mn)$ from equation 2 using Bayes theorem gives;

$$p(Ft, Fn_{0/7}|Mn) = p(Mn|Ft, Fn_{0/7}) \times \left(\frac{p(Ft, Fn_{0/7})}{pMn}\right)$$

Equation 12

Using Equation 11 and Equation 12 the following can be derived;

$$p\left(Ft, Fn_{0/7}|NE\right) = \left(\frac{p\left(Ft, Fn_{0/7}\right) - \left(p\left(Mn|Ft, Fn_{0/7}\right) \times \left(\frac{p\left(Ft, Fn_{0/7}\right)}{pMn}\right) \times pMn\right)}{pNE}\right)$$

This simplifies to;

$$p\left(Ft, Fn_{0/7}|NE\right) = \left(\frac{p\left(Ft, Fn_{0/7}\right) - \left(p\left(Mn|Ft, Fn_{0/7}\right) \times p\left(Ft, Fn_{0/7}\right)\right)}{pNE}\right)$$

From Equation 10;

$$p\left(NE|Ft, Fn_{0/7}\right) = \left(\frac{p\left(Ft,Fn_{0/7}\right) - \left(p\left(Mn|Ft,Fn_{0/7}\right) \times p\left(Ft,Fn_{0/7}\right)\right)}{pNE}\right) * \left(\frac{pNE}{p\left(Ft,Fn_{0/7}\right)}\right)$$

Finally simplifying to;

$$p\left(NE|Ft, Fn_{0/7}\right) = 1 - p\left(Mn|Ft, Fn_{0/7}\right)$$

Equation 13

Note: the proposal to consider conditional probabilities under a Bayesian framework and the original derivation of Equation 13 was from Dr. Ian Overton (IGMM, Edinburgh). The assumptions follow from the Cox-Jaynes axioms[637].

$p(Mn|Ft, Fn_{0/7})$ cannot be calculated from our data as it is unclear which proteins with an $Fn_{0/7}$ are true negative and which are false negatives.

Following assumption (a), all proteins with an Fn>0 are assumed to be present in all normal tissue samples. It is proposed that the probability of a protein being missed in the experiments on normal tissue at a given Ft; $p(Mn|Ft, Fn)$, can be estimated as the proportion of missing quantitation values in the normal samples at the given Ft.

Following this; $p(Mn|Ft, Fn_{0/7})$ can be estimated by $p(Mn|Ft)$. $p(Mn|Ft)$ can be calculated as the average $(1 - Fn)|Ft$ for all proteins where $Fn \neq 0$ (Table 5-6).

Table 5-6 False Negative Probability of Fn for Ft

Ft	pMn|Ft,Fn0	pNE|Ft,Fn0
0	0.8012	0.1988
1/7	0.8095	0.1905
2/7	0.7009	0.2991
3/7	0.6189	0.3811
4/7	0.5147	0.4853
5/7	0.4136	0.5864
6/7	0.2820	0.7180
7/7	0.0846	0.9154

Following these assumptions, it can be concluded that the probability of a protein being a false negative is <10% when Ft = 7/7. Therefore all proteins with $Ft_{7/7}|Fn_{0/7}$ are taken forward and the remainder are excluded from further analysis. In a similar fashion a cut off of $Fn_{7/7}|\ Ft_{0/7}$ was deemed to represent an acceptable probability of being missed (p=0.0778, Table 5-7). This left 2,425 proteins with an expression value for each tissue in at least one experiment or likely to represent a true negative.

Table 5-7 False Negative Probability of Ft for Fn

Fn	pMt|Fn,Ft0	pNE|Fn,Ft0
0/7	0.7723	0.2277
1/7	0.7344	0.2656
2/7	0.5990	0.401
3/7	0.4500	0.55
4/7	0.3855	0.6145
5/7	0.3336	0.6664
6/7	0.2090	0.791
7/7	0.0778	0.9222

For proteins determined to be true negatives an expression value of 10^{-3} was substituted to allow ratios to be calculated. For the remainder, the geometric mean TvO ratios were then determined for each protein, for the PAcIFIC data and TMT data separately. There was a poor correlation between TvO ratios from TMT and PAcIFIC experiments although the intra-experimental correlation was good (Figure 5-9). This is further illustrated by the surprisingly limited relationship between protein abundance assessed by mean TMT reporter ion intensity or spectral count (Figure 5-10).

The method of quantitation appears to have a significant bearing on the final tumour to normal ratio such that the methods cannot be combined. The calculations of pMn and pMt for a given Ft or Fn were therefore reconsidered for the PAcIFIC data alone.

The initial finding of the independence of Ft and Fn from predicted abundance (Figure 5-6A) was re-examined using only the PAcIFIC data (Figure 5-11). This demonstrates that when spectral counts are used, tumour protein abundance was found to correlate with Ft (Pearson's Correlation Coefficient (PCC) =0.6192, $p<0.01$) and normal protein abundance correlates with Fn (PCC = 0.6124).

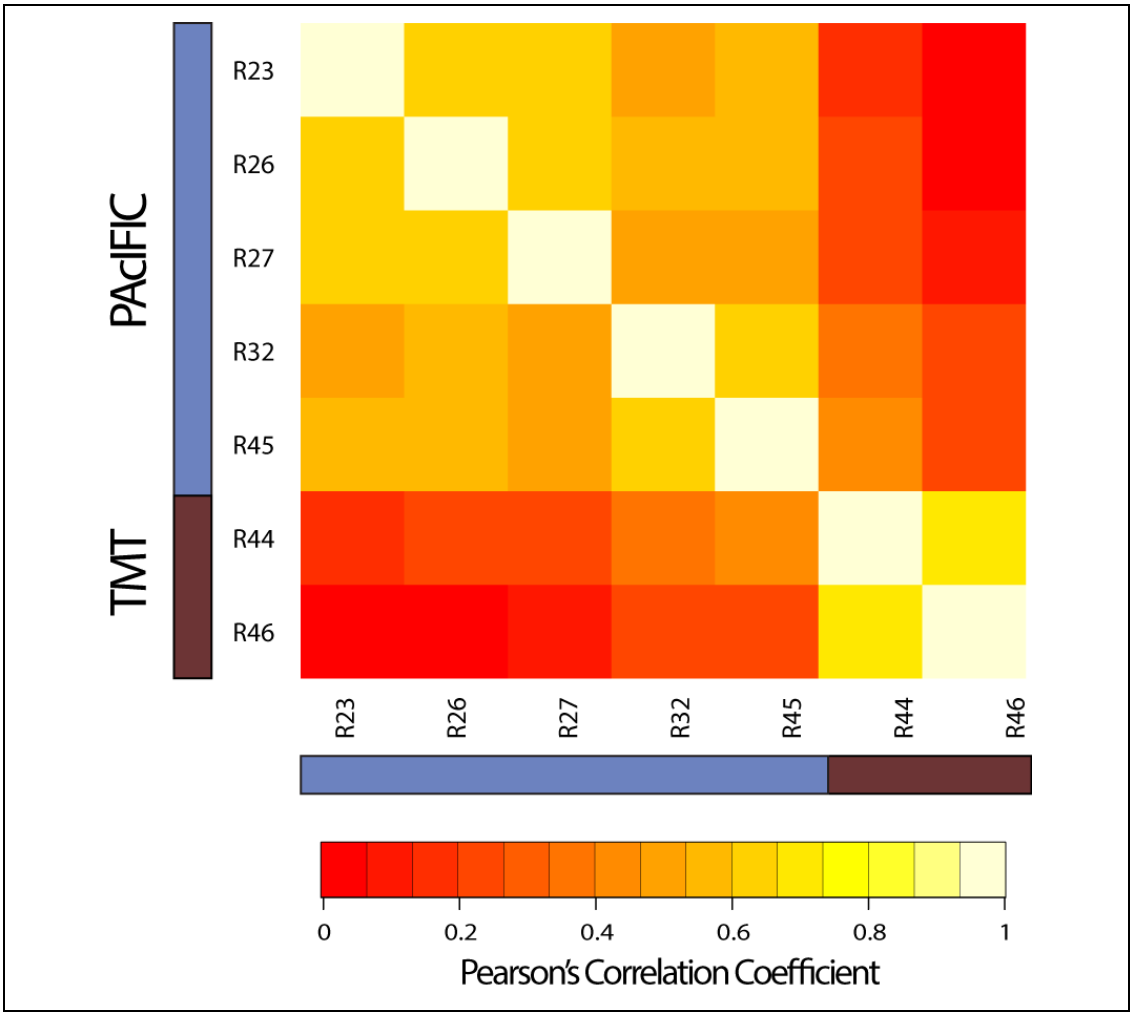

Figure 5-9 Correlation of TMT and PAcIFIC TvO Ratios. R23 = Ratio for patient 23's samples, R26 – Ratio for patient 26's samples etc. The boxes have been shaded according to the Pearson's correlation coefficient between samples (see key).

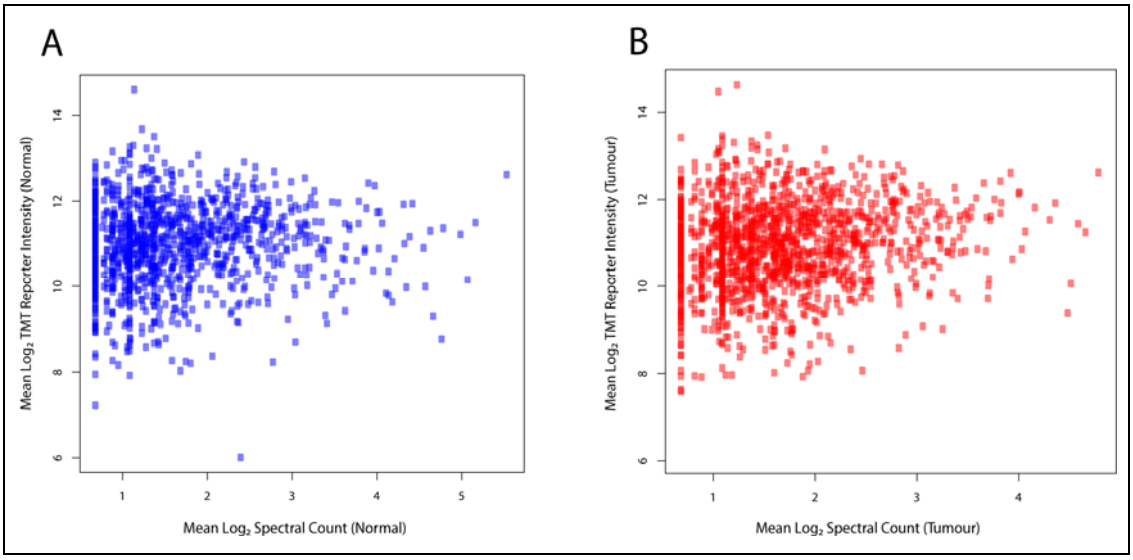

Figure 5-10 Correlation of Spectral Counts and TMT Reporter Ion Intensities for (A) Normal Oesophageal Tissue, (B) OAC Tissue

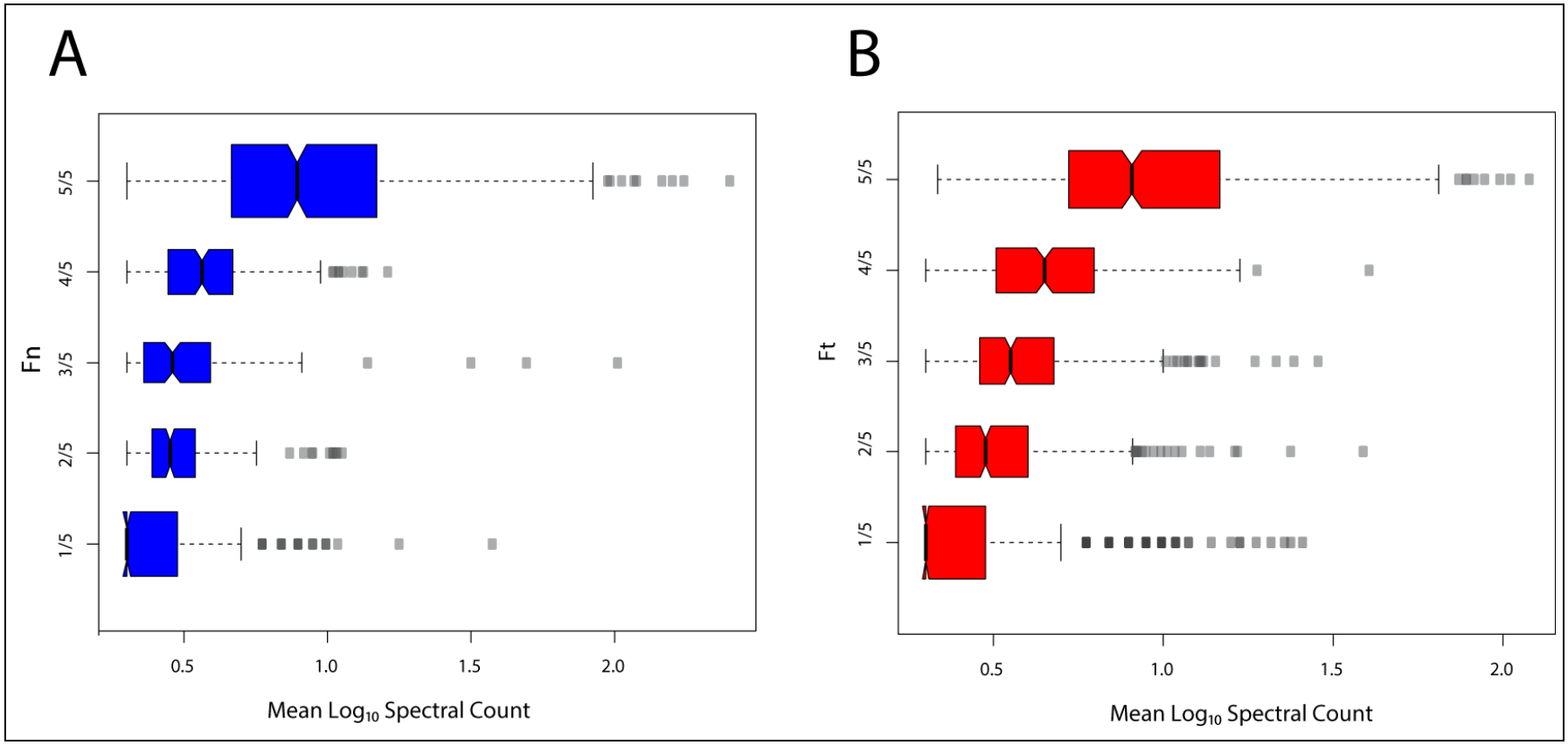

Figure 5-11 Correlation of Spectral Counts and Frequency of Observation in PAcIFIC data for (A) Normal Oesophageal Tissue and (B) OAC Tissue.

5.3.13 False Negative Probability Re-Examined.

The false and true negative probabilities were previously considered using Bayesian statistics to derive Equation 13;

$$p(NE|Ft, Fn_{0/7}) = 1 - p(Mn|Ft, Fn_{0/7})$$

The Bayesian approach has the advantage of allowing liberal assumptions around tumour and normal protein identification probabilities[638].

If protein identifications are considered independent and random events the probability of a false negative can also be evaluated from a frequentist perspective.

For all proteins the total probability (pT) can be considered as;

$$pT = pID + pNoID$$

Where;

pID = probability of a protein being identified;

pNoID = probability of a protein not being identified

Which can be expressed as;

$$pNoID = pNE + pMs$$

Where;

pNE = probability of a protein being not expressed;

pMs = probability of a protein being expressed but missed.

For a single (random) protein the pID from normal tissue can be estimated as Fn.

In the PAcIFIC dataset there are 3 scenarios;

i) Fn = 5/5, ii) Fn = 0/5 and iii) Fn=1-4/5

For i) pT = 1 and Fn = 1 so pNoID must equal zero.

ii) pT=1, Fn = 0 and therefore pNE + pMs = 1 and pNE = 1- pMs

iii) pT=1, and therefore 1= Fn + pNE + pMs.

From assumption (a) we know that pNE must = 0 so therefore;

$$pMs = 1 - Fn$$

Equation 14

The probability of not identifying a protein from a sample is related to technical factors such as peptide biochemical properties, database searching, instrument variables etc. It is therefore predicted that Fn will correlate with Ft for non-tissue specific proteins (i.e. Fn ≠0 and Ft ≠0) fitting assumption (c).

This is found to be true for the PAcIFIC data (PCC = 0.53574).

With this correlation, proteins with the same Ft are predicted to have a similar Mn. This probability can be estimated as the mean (1-Fn) for all proteins for a given Ft. The uncertainty in this probability can be estimated by assuming a Poisson distribution.

The calculated pMn is the probability that a protein at a given Ft will be missed in the normal tissue in one experiment. This is the same (on average) as the number of missed identifications in a series of independent experiments.

From Equation 14 above;

$$pMn|Ft = 1 - (Fn|Ft)$$

Equation 15

The pMn can therefore be calculated for the various levels of Ft.

It is assumed, given the correlation of Ft and Fn, and that experiments are independent and identical, the pMn for 4 nulls when the fifth is detected is the same as the probability for 4 nulls when the fifth is missed. Using this logic, the probability of 5 nulls can simply be estimated as the probability of 1 null[5].

These assumptions can be checked by calculating the predicted probability of 2, 3 and 4 zeroes and the observed probability for various levels of Ft.

The average (1-Fn) was calculated for the set of proteins with Fn≠0.

Average (1-Fn) = pMn = **0.3745**

The converse calculation was performed for the set of proteins where Ft≠0

Average (1-Ft) = pMt = **0.4599**

If experiments are independent and the probability of missing a protein is constant then the predicted and observed probabilities will follow the same distribution (Table 5-8). This is not the case and therefore the average pMn cannot be used to infer the pMn for 5 consecutive experiments. It can be concluded that the Bayesian approach should be preferred for the calculation of pMn or pMt.

Table 5-8 Comparison of Predicted and Actual pMn per Number of Observations

Predicted	Missed n=1	Missed n=2	Missed n=3	Missed n=4	Missed n=5
Normal					
Predicted pMn	**0.3745**	**0.1403**	**0.0525**	**0.0197**	**0.0074**
pMn (actual)	**0.1305**	**0.1077**	**0.1514**	**0.2681**	
Lower Poisson CI 95%	0.1136	0.0925	0.1332	0.2437	
Upper Poisson CI 95%	0.1490	0.1247	0.1714	0.2944	
Tumour					
Predicted pMt	**0.4599**	**0.2115**	**0.0972**	**0.0447**	**0.0205**
pMt (actual)	**0.1298**	**0.1452**	**0.1695**	**0.3426**	
Lower Poisson CI 95%	0.1165	0.1311	0.1542	0.3207	
Upper Poisson CI 95%	0.1440	0.1603	0.1858	0.3656	

The pMn|Ft was calculated using Bayesian methods for the PAcIFIC dataset (Table 5-9). It was estimated that the probability of missing a protein given an Ft of 5/5 is

close to 10% and these proteins could reasonably be considered true negatives. For Ft<5/5 it was unclear if the protein had been missed or was truly not expressed.

High abundance proteins are more likely to be detected and therefore useful data may be gained by selecting high abundance proteins from groups with a lower Ft. The pMn was calculated for the subgroups of proteins with abundances above and below the median geometric mean spectral count for each Ft (Table 5-9). There was a lower pMn with greater abundance at each Ft. The pMn was next calculated for each decile of protein abundance determined from the Log_{10} geometric mean spectral count, and for all proteins above an abundance threshold (Table 5-10).

With each increasing decile of tumour protein abundance, the pMn decreased. The pMn was plotted for incrementing thresholds of tumour protein abundance (Figure 5-12). From this curve, a cut-off representing a pMn of 10% was determined as a Log_{10} Tumour spectral count > 0.905 (Geometric mean spectral count >8.1). This corresponds to 47 proteins (3.6%) in the set $Ft|Fn_{0/5}$ (total n=1302 proteins). In comparison, the set of $Ft_{5/5}$ $|Fn_{0/5}$ only contains 34 proteins which, using the Bayesian model, have an estimated pMn of 12%.

The effect of combining abundance and Ft was next examined to identify if the number of proteins selected with a pMn <10% could be increased. Each Ft was then subdivided according to percentile abundance and pMn calculated (Table 5-11).

Within each Ft category a threshold was iterated above which the observed pMn <10%. The number of proteins in the set $Ft|Fn_{0/5}$ for each of these groups was then calculated, giving a total of 24 (Table 5-12).

From these analyses, the greatest number of proteins with a pMn<10% can be selected by using an abundance criteria alone.

To determine if the same holds true for pMt, the same analysis was performed for deciles of normal protein abundance (Table 13). This demonstrated an abundance threshold around the 90^{th} percentile defined proteins with a pMt of <10%. In the set of $Fn|Ft_{0/5}$ this identifies 8 proteins (3%) as being true negatives. Using the Bayesian

model, seven proteins were identified in the set $Fn_{5/5}|Ft_{0/5}$ with an estimated pMt of 12%. There were only 32 proteins in the combined set $Fn_{3/5\text{-}5/5}|Fn_{0/5}$ negating any benefit in subdividing categories by abundance. So similar to pMn|Ft, using an abundance threshold alone is the preferred criteria for identifying normal specific proteins.

Table 5-9 pMn|Ft for the PAcIFIC Data

Ft	pMn|Ft	n (Proteins)	Lower 95% CI	Upper 95% CI	pMn|Ft lower 50% Abundance	pMn|Ft upper 50% Abundance
0/5	**0.6893**	253	0.6438	0.7369	**NA**	**NA**
1/5	**0.5860**	157	0.5332	0.6422	**0.6509**	**0.5510**
2/5	**0.5023**	213	0.4607	0.5471	**0.5385**	**0.4679**
3/5	**0.4292**	226	0.3914	0.4695	**0.4874**	**0.3645**
4/5	**0.3235**	277	0.2940	0.3552	**0.3842**	**0.2623**
5/5	**0.1207**	545	0.1079	0.1346	**0.1949**	**0.0463**

Table 5-10 pMn for Deciles of Abundance and for All Proteins Above an Abundance Threshold

Proteins within Decile of Tumour Abundance	pMn	Proteins Above Threshold - Mean Log_{10} TSC (percentile)	pMn
0-10%	**0.55**	All	**0.3183**
10-20%	**0.5**	>0.389 (10%)	**0.2976**
20-30%	**0.4840**	>0.477 (20%)	**0.2779**
30-40%	**0.4169**	>0.539 (30%)	**0.2437**
40-50%	**0.3603**	>0.619 (40%)	**0.2101**
50-60%	**0.3338**	>0.692 (50%)	**0.1803**
60-70%	**0.2350**	>0.759 (60%)	**0.1420**
70-80%	**0.1915**	>0.853 (70%)	**0.1107**
80-90%	**0.1103**	>0.976 (80%)	**0.0704**
90-100%	**0.0288**	>1.195 (90%)	**0.0288**

TSC - Tumour Spectral Count

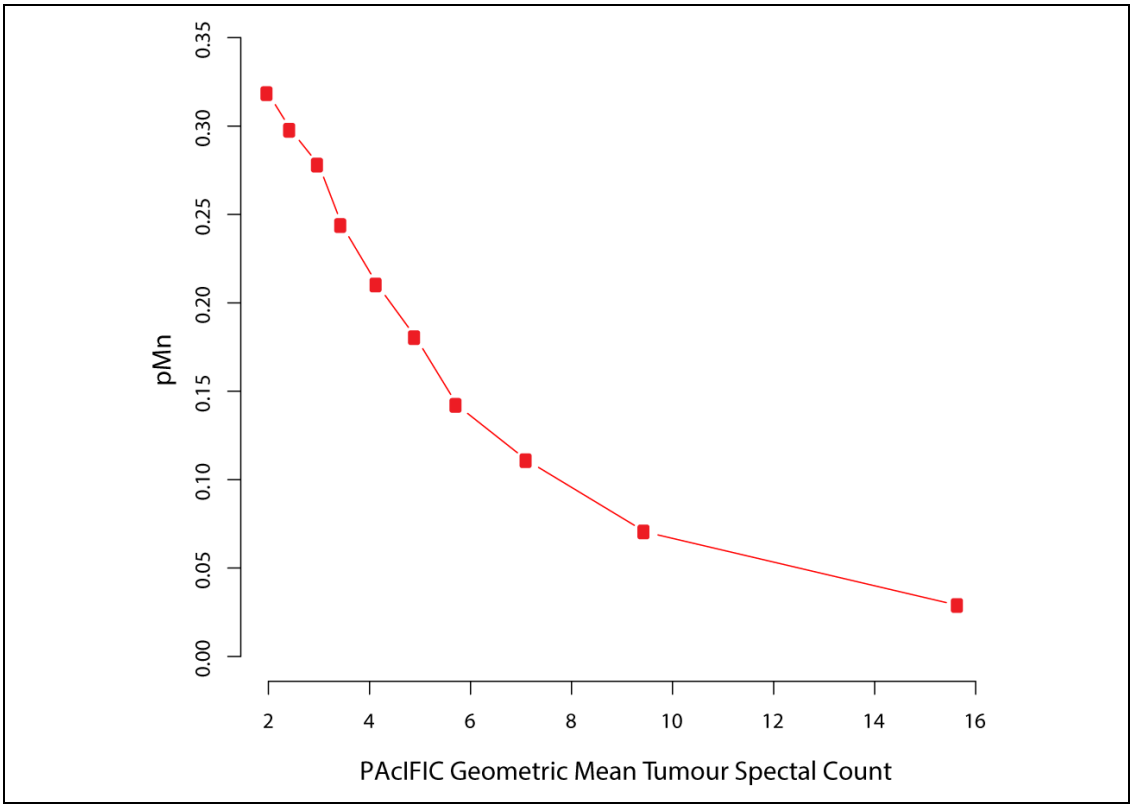

Figure 5-12 The Effect of PAcIFIC Tumour Spectral Count on pMn.

Table 5-11 pMn|Ft for Geometric Mean Spectral Count Percentiles

Ft	Geometric Mean Tumour Spectral Count (Percentile)											
	>10%	>15%	>20%	>40%	>50%	>60%	>70%	>80%	>90%	>95%	>99%	>99.9%
1/5				0.5510	0.5510	0.5510	0.5461	0.56	0.5417	0.6	0.5333	0.8
2/5				0.4906	0.4679	0.4419	0.4412	0.4314	0.4381	0.4727	0.2667	0.4
3/5				0.3773	0.3649	0.3596	0.3281	0.2875	0.25	0.2	0	0
4/5				0.2683	0.2596	0.2288	0.2048	0.1893	0.15	0.157	0	0
5/5	0.1014	0.094	0.0825	0.0556	0.0463	0.0348	0.0171	0.0183	0.0110	0	0	0

Table 5-12 Number of Proteins Identified with pMn <10% using combined Ft and Abundance Thresholds

Ft	Threshold of Mean Log_{10} TSC	pMn	Number of Proteins for Ft|Fn0
5/5	>0.61	0.0994	**22**
4/5	>1.165	0.0667	**0**
3/5	>1.124	0	**2**

TSC – Tumour Spectral Count

Table 5-13 pMt for All Proteins above an Abundance Thresholds

Mean Log_{10} Normal Spectral Count (percentile)	pMt
All	0.2816
>0.0 (40%)	0.2816
>0.3010 (50%)	0.2816
>0.3890 (60%)	0.2373
>0.5000 (70%)	0.1838
>0.6571 (80%)	0.1540
>0.9624 (90%)	0.1037

The predicted tumour and normal-specific proteins are detailed in Table 5-14 and Table 5-15. These proteins were added to the data set Ft ≠ 0, Fn ≠ 0 to generate a final complete set of PAcIFIC quantified proteins (n=1473). Several of these proteins had been identified in the initial TMT study as highly expressed in tumour tissue compared to normal oesophagus, increasing confidence in this selection method (see discussion).

Table 5-14 Predicted Tumour Specific Proteins from PAcIFIC data.

Uniprot ID	Protein name	Primary Subcellular location
ACDSB_HUMAN	Short/branched chain specific acyl-CoA dehydrogenase,	Mitochondria
ACSL5_HUMAN	Long-chain-fatty-acid--CoA ligase 5	Mitochondria
AGR2_HUMAN	Anterior gradient protein 2 homolog	Secreted
AIFM1_HUMAN	Apoptosis-inducing factor 1	Mitochondria
ANX11_HUMAN	Annexin A11	Cytoplasm
CO4A_HUMAN	Complement C4-A	Secreted
EF1G_HUMAN	Elongation factor 1-gamma	
F16P1_HUMAN	Fructose-1,6-bisphosphatase 1	
H13_HUMAN	Histone H1.3	Nucleus
HCDH_HUMAN	Hydroxyacyl-coenzyme A dehydrogenase	Mitochondria
HUWE1_HUMAN	E3 ubiquitin-protein ligase HUWE1	Cytoplasm
HV319_HUMAN	Ig heavy chain V-III region JON	
IGHG1_HUMAN	Ig gamma-1 chain C region	Secreted
ITIH2_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H2	Secreted
K1C20_HUMAN	Keratin, type I cytoskeletal 20	Cytoplasm
LEG2_HUMAN	Galectin-2	
LEG4_HUMAN	Galectin-4	
MUC2_HUMAN	Mucin-2	Secreted
MUC5A_HUMAN	Mucin-5AC	Secreted
MUC6_HUMAN	Mucin-6	Secreted
NOMO2_HUMAN	Nodal modulator 2	ER membrane
OLFM4_HUMAN	Olfactomedin-4	Secreted
PERE_HUMAN	Eosinophil peroxidase	Cytoplasmic granule
PIGR_HUMAN	Polymeric immunoglobulin receptor	Cell membrane
ROA3_HUMAN	Heterogeneous nuclear ribonucleoprotein A3	Nucleus
SUIS_HUMAN	Sucrase-isomaltase, intestinal	Cell membrane
SYVC_HUMAN	Valine--tRNA ligase	
TOP2A_HUMAN	DNA topoisomerase 2-alpha	Cytoplasm
VILI_HUMAN	Villin-1	Cytoplasm
B3KM80_HUMAN	Nucleolin, isoform CRA_c	
Q5TB52_HUMAN	3'-phosphoadenosine 5'-phosphosulfate synthase 2	
Q6FHB6_HUMAN	Annexin	
B7Z8A9_HUMAN	cDNA FLJ51166, highly similar to Matrix metalloproteinase-9	
Q6FH49_HUMAN	Nicotinamide N-methyltransferase protein	
Q9HBC6_HUMAN	Intestinal mucin 2	
Q6IB91_HUMAN	PCK2 protein	
Q8NC03_HUMAN	cDNA FLJ90630 fis, clone PLACE1003438	
A8KA66_HUMAN	cDNA FLJ78485, highly similar to Homo sapiens epidermal growth factor receptor pathway substrate 8	
Q5R209_HUMAN	Carbamoylphosphate synthetase I	
Q6I9V1_HUMAN	DPEP1 protein	
Q3KT79_HUMAN	Alpha-methyl-acyl-CoA racemase	
Q53SB2_HUMAN	Putative uncharacterized protein RPIA (Fragment)	
B4DE36_HUMAN	Glucose-6-phosphate isomerase	Cytoplasm
B3KMX0_HUMAN	cDNA FLJ12837 fis, clone NT2RP2003228, highly similar to DNA replication licensing factor MCM4	
E9PH38_HUMAN	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	
B2R778_HUMAN	cDNA, FLJ93317, highly similar to Homo sapiens cadherin 17	
Q5U0I7_HUMAN	Serine/threonine-protein phosphatase	

Protein names and subcellular location are provided from Uniprot Annotations.

Table 5-15 Predicted Normal Specific Proteins from PAcIFIC data.

Entry name	Protein names	Primary Subcellular location
ACTB_HUMAN	Actin, cytoplasmic 1 (Beta-actin)	Cytoplasm
ARP2_HUMAN	Actin-related protein 2 (Actin-like protein 2)	Cytoplasm
CMA1_HUMAN	Chymase (Alpha-chymase) (Mast cell protease I)	Secreted
DSG1_HUMAN	Desmoglein-1 (Cadherin family member 4)	Single-pass type I membrane protein.
LMOD1_HUMAN	Leiomodin-1 (64 kDa autoantigen 1D)	Cytoplasm
A1L0V1_HUMAN	ACTN1 protein (Fragment)	
B7Z797_HUMAN	cDNA FLJ56799, highly similar to Sciellin	
Q6LD69_HUMAN	Elafin (Fragment)	

Protein names and subcellular location are provided from Uniprot Annotations.

The Log_2 Tumour: Normal ratios were again calculated for TMT and PAcIFIC experiments independently with a quantitation value of 10^{-3} substituted for predicted true negative proteins. The correlation between TMT and PAcIFIC TvO ratios remained poor (Figure 5-13).

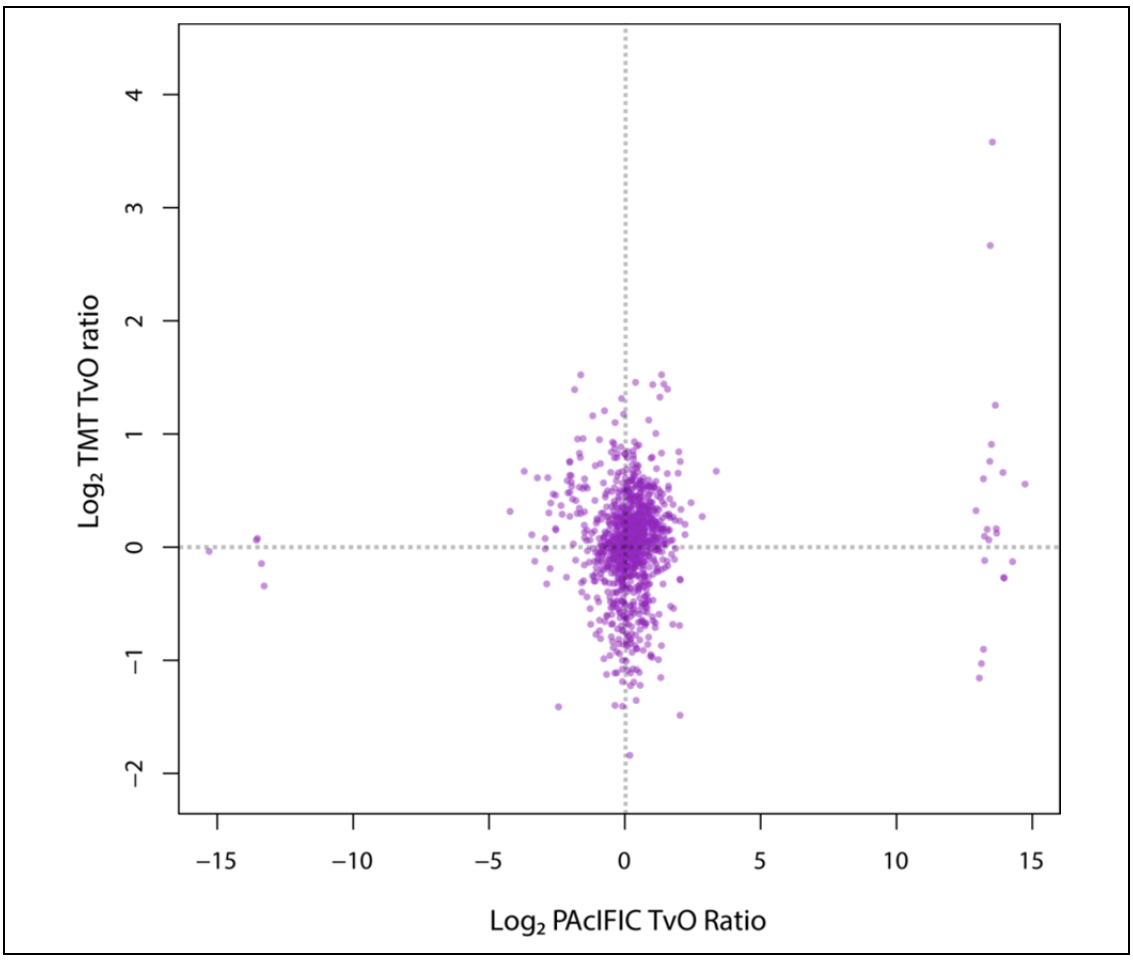

Figure 5-13 Correlation of TMT and PAcIFIC TvO Ratios. Proteins quantified in both experiments are plotted according to their Log_2 TvO ratios. Nominally tumour or oesophageal specific proteins from the PAcIFIC data have been included.

5.3.14 Comparison of Spectral Counting and Isobaric Quantitation

To further assess the relationship of spectral counts and TMT reporter ion intensities, spectral counts were extracted for the peptide-spectrum matching files from the mFASP TMT experiments. To maximise the number of quantifiable peptides, a liberal protein identification threshold was used with a peptide false discovery rate of <5% and 1 unique peptide per protein accepted. Spectral counts were available for 3,310 proteins identified in experiment p44 and 2,735 proteins in experiment p46.

The mean $\log_{10}$ gastric reporter ion intensity was calculated for each protein and then compared to the $\log_{10}$ spectral count on a per protein basis. Only a weak correlation was observed between spectral counts and mean reporter ion intensities derived from the same peptide-spectrum matching file (Figure 5-14). The limited correlation may be secondary to differences in the quantitative distributions (Figure 5-15). Spectral counts show a reciprocal distribution with most proteins demonstrating a spectral count of 1. In contrast, the TMT reporter ion intensities approach a Gaussian distribution.

The correlation between patients using the same quantitation method was also assessed. This revealed a significantly greater concordance (Figure 5-16). This suggests both methods are internally consistent but not comparable. For subsequent analyses, PAcIFIC and TMT datasets were therefore considered independently.

The TvO ratios from both datasets showed Gaussian distributions and were median-normalised. As different dynamic ranges were observed between TMT and PAcIFIC data, both datasets were normalised to generate z-scores (dividing all ratios in each experiment by the experimental standard deviation so the experimental distributions become equivalent). The complete list of 2,455 quantified proteins and the z-scores for each experiment are provided in the Appendix.

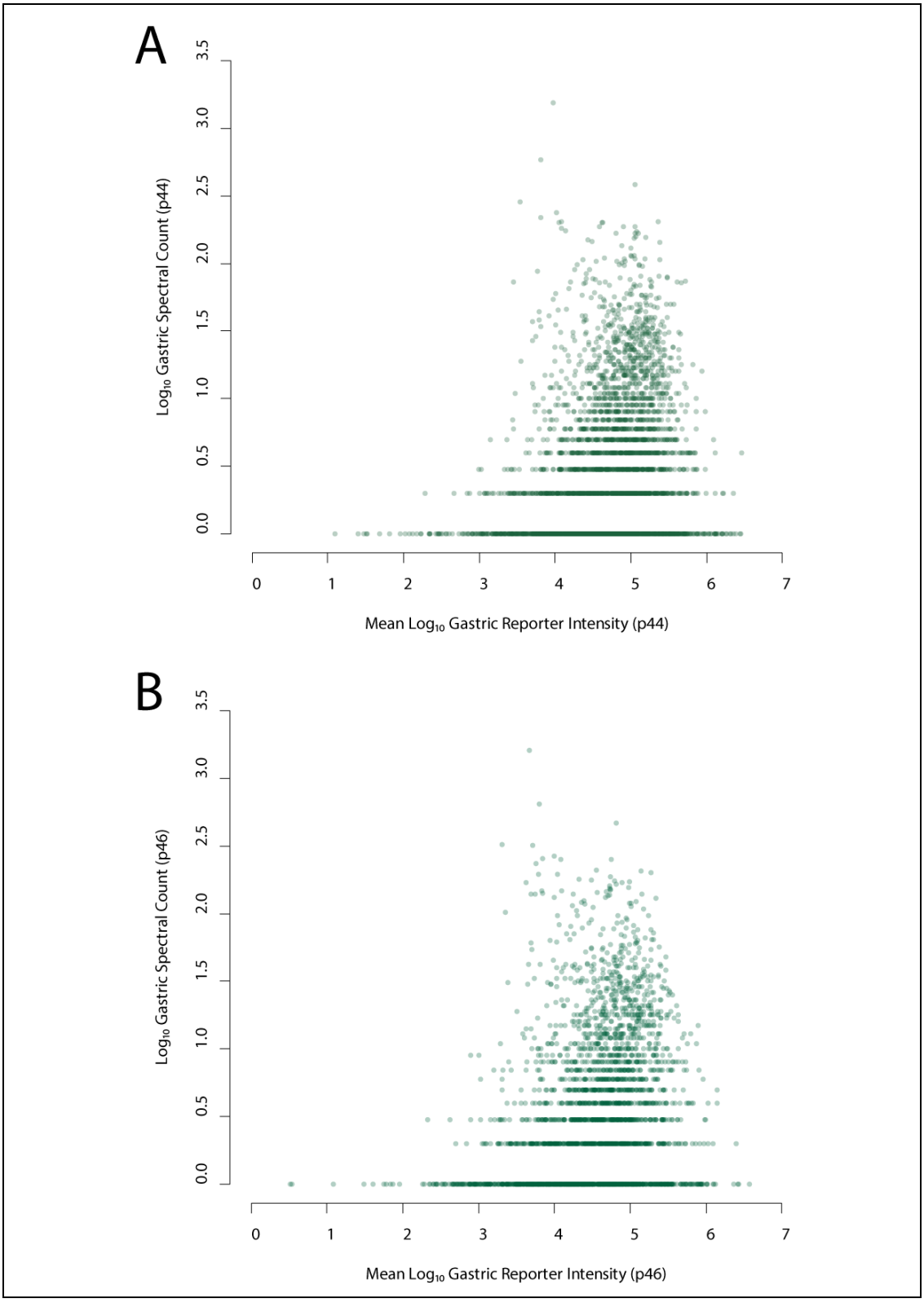

Figure 5-14 Correlation of TMT Reporter Intensity and Spectral Count for Patient 44 (p44) and Patient 46 (p46). Proteins are plotted as points according to their log_{10} spectral count and mean log_{10} TMT reporter intensity for gastric tissue. Spearman correlation coefficients; p44 = 0.215 ($p<0.01$); p46 = 0.255 ($p<0.01$).

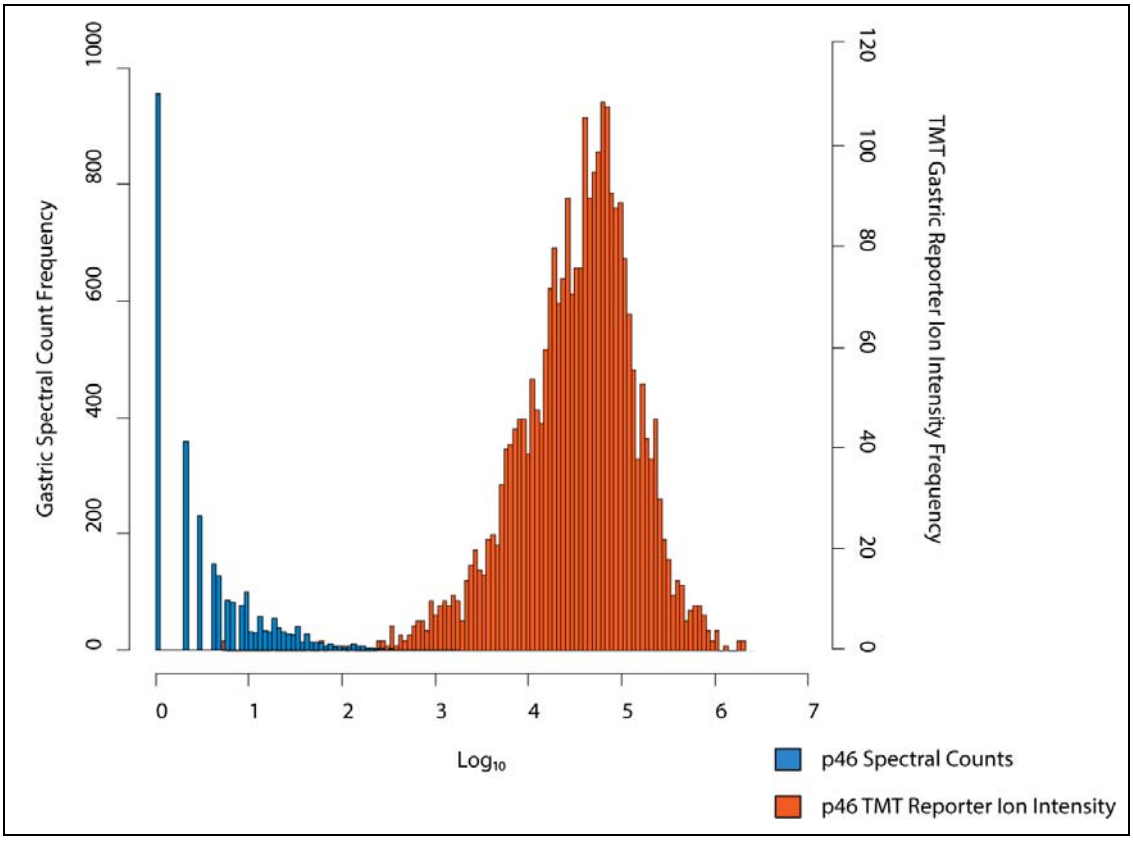

Figure 5-15 Distributions of Spectral Counts and TMT Reporter Intensities for Patient 46 (p46). Frequency histograms are shown according to the log_{10} spectral count or log_{10} reporter ion intensity. Data representative of patient 44.

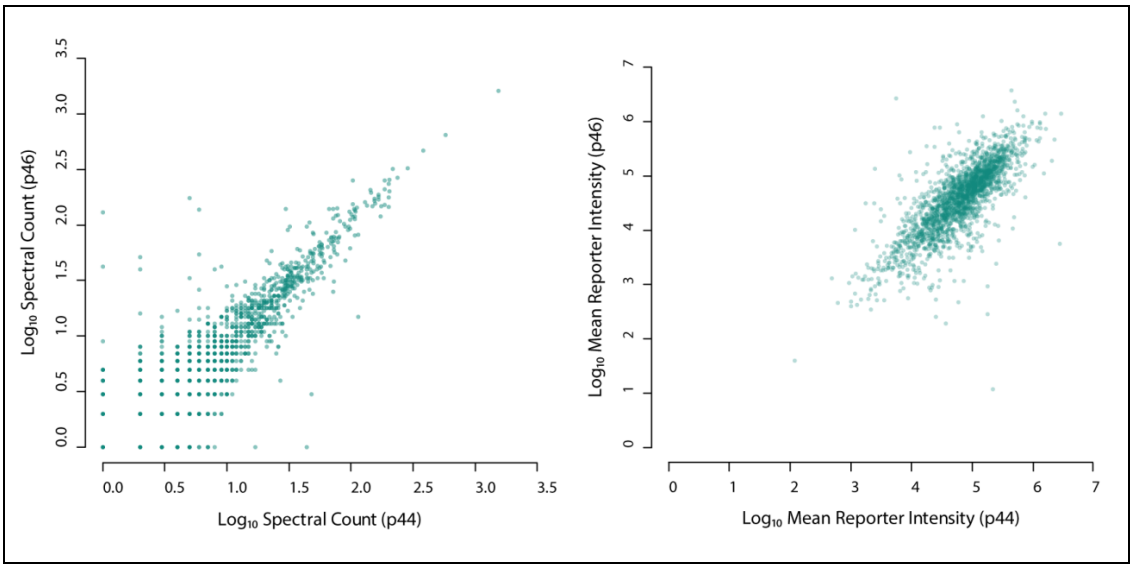

Figure 5-16 Correlation of Spectral Counts and Gastric TMT reporter Intensities for Patient 44 (p44) and Patient 46 (p46). Proteins quantified in both patients were plotted according to their log_{10} spectral count or mean log_{10} TMT reporter intensity for gastric tissue. Spearman correlation coefficients; spectral counts = 0.8753 ($p<0.0001$), TMT reporter intensities = 0.7869 ($p<0.0001$).

5.4 Discussion

Following the successful application of quantitative shotgun proteomics to oesophageal tissue in the previous chapter, attempts were made to enhance the number of protein identifications. The combination of doubling the number of OFFGEL fractions and applying the mFASP tissue lysis method led to an increase in proteome coverage. The increase in identifications was also associated with an increase in the proportion of membrane proteins identified. This likely due to the use of a higher detergent concentration in the lysis buffer, enhancing the solubilisation of hydrophobic proteins.

The published FASP method describes on-column tryptic digestion which has the advantage of minimising sample transfer and potentially reducing protein loss. In this study, off-column digestion appeared more efficient. Failure of tryptic digestion using the on-column method could explain retention of the majority of the protein mass in the NMWCO filters. When the identical protocol was applied off-column, however, the tryptic digestion appeared near-complete. Failure of filtration of the tryptic peptides after on-column digestion appears to be the most likely explanation.

A further surprising feature of the mFASP protocol was the preservation of proteins with an apparent molecular weight <30kDa despite the use of 30kDa NMWCO filters. This finding has been previously demonstrated and the reason is not yet clear although it may reflect post-translational protein modifications including glycosylation and phosphorylation that restrict the passage through the filter[633]. An alternative is that partial protein folding or aggregation is achieved in solution and this prevents filtration. The presence of detergent, the reducing agent DTT and high urea concentration would, however, be expected to maintain proteins in a denatured state. Conversely, SDS contamination was not observed in the mass spectra suggesting this had been effectively filtered from the lysates.

The mFASP TMT method appeared to enhance both proteome coverage and quantitative dynamic range compared to PAcIFIC. This was unexpected as the TMT experiments required 30 hours of mass spectrometry time to analyse three samples from the same patient. In contrast a single sample was analysed over 24 hours with

the PAcIFIC method. The increased data acquisition time and data-independent work-flow in the PAcIFIC method has previously been reported to enhance proteome coverage[404,406,639]. The superior performance of TMT may reflect the beneficial reduction in sample complexity produced by fractionation. In support of this, the combination of strong-cation exchange fractionation and high mass accuracy Orbitrap mass spectrometry has recently been reported to yield over 7,000 protein identifications from a single laser capture micro-dissected sample[640].

To enhance the power of subsequent analyses, an attempt was made to combine all the proteomic datasets. Tumour to normal oesophageal expression ratios allowed quantitative data obtained via different approaches to be combined. Missing data from the PAcIFIC experiments limited the application of this method and represents a fundamental problem in proteomic informatics. A Bayesian statistical approach, developed in collaboration with Dr. Ian Overton, was found to offer superior performance over a frequentist strategy in the estimation of the probability a protein was not expressed in a tissue.

Bayesian statistical methods have become widely used in recent years with a myriad of applications across biological, environmental, computer and even political sciences[641]. A method to deal with missing proteomic data from 2D gels has previously been developed. In this study a likelihood-based approach out-performed parametric statistics applied using a frequentist framework[642]. A further study addressing the same problem also supported the use of a likelihood estimate[643]. In that work, protein abundance was determined from spot volume on 2D gels and was correlated with the frequency of spot detection in a series of 2D gels. This supports the finding of a correlation between abundance and Ft or Fn in this study.

The Bayesian method was applied to the PAcIFIC dataset to derive a threshold of tumour protein abundance (Ft) at which the absence of normal protein detection was thought to be unlikely ($p<0.1$) to represent a missed identification. The reverse was also applied for normal protein abundance and the tissue specific proteins (Table 5-14, Table 5-15) added to the dataset of proteins with at least one identification in each tissue type.

Several candidate tumour specific proteins (Anterior Gradient 2, Polymeric Immunoglobulin receptor, Mucin-2) have been previously independently identified as highly expressed in OAC compared to normal oesophagus in the pilot TMT study (Table 4-2). Other candidates (Villin, Sucrase Isomaltase) have been previously reported to be specific for intestinal type epithelium and highly expressed in Barrett's and OAC[644,645]. This supports the validity of the method to identify truly tissue specific proteins. Evaluation of the novel candidates by IHC would allow the accuracy of this approach to be fully determined.

A poor correlation was observed between spectral counts from the PAcIFIC experiments and mean reporter ion intensities from the TMT experiments. This may reflect fundamental differences between spectral counting and isobaric methods of quantitation. In particular, spectral counting has only moderate reproducibility between technical replicates and has significantly lower accuracy than isobaric methods[437].

To address this, spectral counts and reporter ion intensities were compared from the same peptide-spectrum matches. This revealed a persistently poor correlation between methods. Interestingly the correlation was significantly improved when the same method was compared across patient data sets. This suggests a lack of compatibility between these quantitative methods. It was concludes that the datasets should therefore be treated separately.

One possibility for the poor correlation between PAcIFIC and TMT TvO ratios lies in the method used for TMT analysis. To obtain a summary TvO ratio for a protein, the geometric mean of all tumour peptide reporter ion intensities was calculated and then the geometric mean of all normal oesophageal peptide reporter ion intensities before the TvO ratio was derived. For most proteins, tryptic digestion would be expected to produce peptides in a one to one ratio and therefore, if all peptides are detected, reporter ion intensities generated by each peptide from the same protein should be roughly equivalent. However, some peptides are more easily ionisable and are therefore preferentially detected, leading to higher reporter ion intensities. The geometric mean may not be the best summary representation of all the reporter ion

intensities from one protein if the distribution is heavily skewed. Spectral counts, although still subject to significant limitations as a quantitative method, would not be influenced by fragment ion intensity.

The ratio of reporter ion intensities for each peptide relies only on the starting amount of peptide from each tissue type labelled and, as the same peptide is quantified from each tissue type, this ratio is independent of any bias in peptide detection probability. For this reason, in subsequent TMT analyses (Chapter 7), the ratio of reporter intensities was first calculated for each peptide and then the geometric mean of all peptide ratios calculated. This normalises for potential differences in fragment ion detection.

In conclusion, dysregulated proteins identified from these experiments may be targets for immunotherapy and could be validated individually for tissue-specific expression by IHC. This would require an extensive effort and tissue resource to complete. If a targetable oncogenic pathway is upregulated in OAC, several members of that pathway may be dysregulated. Unfortunately, incomplete proteomic datasets make the identification of dysregulated pathways a challenge. One approach to tackle this involves network analysis and will be examined in the next chapter.

Chapter 6: Network Analysis of Quantitative Proteomic Data

6.1 Chapter Aims

a) To define functional clusters in a human functional linkage network.

b) To define groups of dysregulated proteins from PAcIFIC and TMT datasets and map them to functional clusters.

c) To identify enriched functional clusters as evidence of enriched biological functions in the dysregulated protein lists.

d) To visualise the functional clusters in a network and identify targetable pathways.

6.2 Introduction

Quantitative protein expression studies from human tissue have the potential for greater functional relevance compared to microarray or next-generation sequencing derived transcriptomic data. This advantage may be offset by the difficulties in interpreting the incomplete datasets generated. Dysregulated proteins can be annotated and considered on an individual candidate basis as performed in the initial TMT study. This labour-intensive approach is untenable for large datasets.

Network analysis offers the possibility to identify components of functional pathways from lists of dysregulated proteins. This allows candidate dysregulated pathways to be selected for more detailed investigation using hypothesis-directed approaches such as those applied in Chapter 3. The method is based on the same principle underlying gene set enrichment analysis (GSEA); the technique where genes are annotated with a functional ontology and enriched ontologies are identified from gene lists[646].

Despite the frequent use of GSEA in cancer proteomic studies, novel therapeutic targets are rarely identified as the ontologies often describe generic cellular processes e.g. cell cycle regulation, rather than druggable pathways and the pathway context is not appreciated. By using a functional linkage network as the background rather than gene ontologies, enriched functional modules can be identified by "seeding" the network using a list of genes. This approach is similar to the published gene-network enrichment strategy[647,648]. The network connections can also be preserved with this method, allowing potential regulatory interactions to be revealed. This approach will be employed in this chapter using the quantitative proteomic data derived from PAcIFIC and mFASP TMT experiments.

6.2.1 A Human Functional Linkage Network for Biological Process Discovery

The largest published human functional linkage network (FLN), referred to as the Linghu network[460], will be used for the analysis in this chapter. The edges of this network were constructed using Bayesian statistics to incorporate and weigh the evidence for a functional link across 16 different domains (Table 6-1). The complete

network consists of 21,657 nodes (genes) and 22,388,609 edges with a likelihood ratio of greater than 1 (implying greater evidence for than against a functional linkage).

Table 6-1 Data sources used to construct the Linghu FLN. Adapted from reference[460]

Data source	Description	Number of unique gene pairs	Number of unique genes
Curated PPI	Curated human PPI from HPRD, BIND, BioGRID, IntAct, MIPS, DIPS, and MINT[649-655]	90,352	10,281
Y2H	PPI from high-throughput yeast two-hybrid experiments[656]	2,611	1,522
MS	PPI from large-scale mass spectrometry experiments[657]	2,046	1,159
DDI	Protein pairs containing interacting protein domains[658,659]	6,933,469	13,454
Co-expression	Expression correlation from multiple large-scale expression datasets[660-663]	5,110,798	16,287
Domain Sharing	Proteins pairs sharing same protein domains[664]	2,064,262	17,328
Phylogenetic	Gene pairs having correlated phylogenetic profiles[665]	18,086	2,607
Gene Loci	Gene pairs located close to each other along the chromosome[660]	10,070	1,365
Fusion	Protein pairs fused into one single protein in other species[660]	361	361
Yeast	Functional associations mapped from seven types of functional genomics data in yeast through gene orthology[665]	123,380	3,809
Worm	Functional associations mapped from four types of functional genomics data in worm through gene orthology[666]	96,911	5,737
Fly	Functional associations mapped from three types of fly functional genomics data through gene orthology[660]	139,984	5,966
Mouse-rat	Functional associations mapped from three types of rodent functional genomics data through gene orthology[660]	254,477	11,789
Text Mining	Co-occurrence in PubMed abstracts[660]	518,716	12,286
Molecular Function	Gene pairs sharing same molecular function terms in GO[667]	6,937,725	7,863
Cellular Component	Gene pairs sharing same cellular component terms in GO[667]	5,591,796	12,503

Abbreviations. PPI – protein-protein interaction, HPRD - Human Protein Reference Database, BIND - Biomolecular Interaction Network Database, BioGRID - Biological General Repository for Interaction Datasets, IntAct – European Bioinformatics Institute Molecular Interaction Database, MIPS - Mammalian Protein-Protein Interaction Database, DIPS – Database of Interacting Proteins, MINT – Molecular Interaction Database, Y2H – Yeast two hybrid assay, DDI – Domain-Domain Interactions, GO – Gene Ontology.

6.2.2 Identifying Functional Clusters

The Markov clustering algorithm, MCL can be used to uncover functional groups (clusters) within biological networks[523]. This unsupervised clustering method has been widely applied in bioinformatics and has been comprehensively reviewed[668]. Briefly, this algorithm uses a random walk strategy to identify groups of genes with high connectivity. Within a functional linkage network such as the Linghu network, these groups contain genes defined by similar functional properties and can be considered as functional clusters. The size of clusters identified by MCL can be determined by a user-definable inflation value.

6.3 Results

6.3.1 Generating Functional Clusters in the Linghu Network.

The original Linghu network was designed to be inclusive and contains many edges with relatively low levels of biological evidence. The network was therefore filtered at 2 more stringent levels of likelihood ratio (LR); LR $\geq$ 5 and LR $\geq$ 20 (LR5+, LR20+) to ensure edges (functional interactions) had high or very high levels of evidence. These filtered networks were then subjected to Markov clustering using the MCL algorithm[668] to generate clusters of genes with similar biological functions. This was performed with inflation (I) values from 2.0 to 6.0 to generate clusters of varying size. The effect of varying the MCL I-value on the number of clusters in both filtered networks is demonstrated in Figure 6-1.

6.3.2 Defining Differentially Expressed Genes from Proteomic Datasets

The Uniprot Accessions for proteins identified from the PAcIFIC and mFASP TMT experiments (protein IDs) were mapped to Entrez GeneIDs and searched against the genes in the LR5 and LR20 Linghu networks (for details see Methods 2.11.4). The mapped genes were used for subsequent analyses. For the PAcIFIC and mFASP TMT lists, false discovery rate (FDR) corrected p-values were calculated for each protein z-score using the Benjamini Hochberg method[526] (details of R commands provided in Appendix B). Lists of proteins with an FDR-corrected $p<0.1$ and FDR-corrected $p<0.05$ were generated. As an alternative strategy, the magnitude of

relative TvO expression was used deferring statistical confidence estimation to the network analysis stage. Thresholds of either 1.5 fold or 2 fold dysregulated were selected and all proteins outside these thresholds included, irrespective of the FDR-corrected p-value.

Differentially expressed gene lists and corresponding random resamples containing the same number of genes were then mapped to the clusters generated from the LR5+ and LR20+ FLNs at each I-value (details of Perl scripts and R commands provided in Appendix B). FDR-corrected p-values were calculated for each cluster using a previous described method[462].

6.3.3 Optimising Network Likelihood Ratio and Cluster Inflation Value.

To determine the optimal I-value for the identification of enriched clusters, two thresholds were used; an FDR-corrected p-value <0.1 and an arbitrary threshold incorporating only clusters with greater than 20 genes and >10% of the genes mapped. The number of clusters fulfilling these criteria for both TMT and PAcIFIC experiments, across the range of I-values and dysregulated protein lists, was calculated (Table 6-2).

The LR20+ network provided very limited coverage of the identified proteome and was not considered any further. The clusters from I-values above 5.0 were discarded from the LR5+ network for the same reason. The remaining results across I-values, and differentially expressed thresholds are displayed in Table 6-2. An FDR-corrected p<0.1 allowed more clusters to be identified at each I value than the arbitrary cut-off.

Network complexity and visual information delivery are inversely related. The cluster connectivity and size varied significantly with the I-value. Network visualisations were therefore reviewed and an optimal threshold iterated for each example where significantly enriched clusters were identified, balancing enriched cluster number, size and connectivity. The optimal range for visual information delivery for the TMT experiments was an I-value between 2 and 3 using the 1.5 fold

up and down regulated gene lists. For the PAcIFIC dataset the optimal range was an I-value of between 2 and 2.5 with the 2 fold up and down regulated lists. Visualisations of these networks were created in the open-source Cytoscape environment[647] using force-directed and spring-embedded approaches to produce the optimal grouping of functional clusters (Figure 6-2 to Figure 6-5).

For each of the networks, the cluster function was summarised using DAVID biological process gene ontologies[669]. The low FDR-corrected p-values for DAVID gene ontologies emphasises the relative enrichment of functionally similar genes within Markov clusters generated from the Linghu FLN (Table 6-3 and Table 6-4).

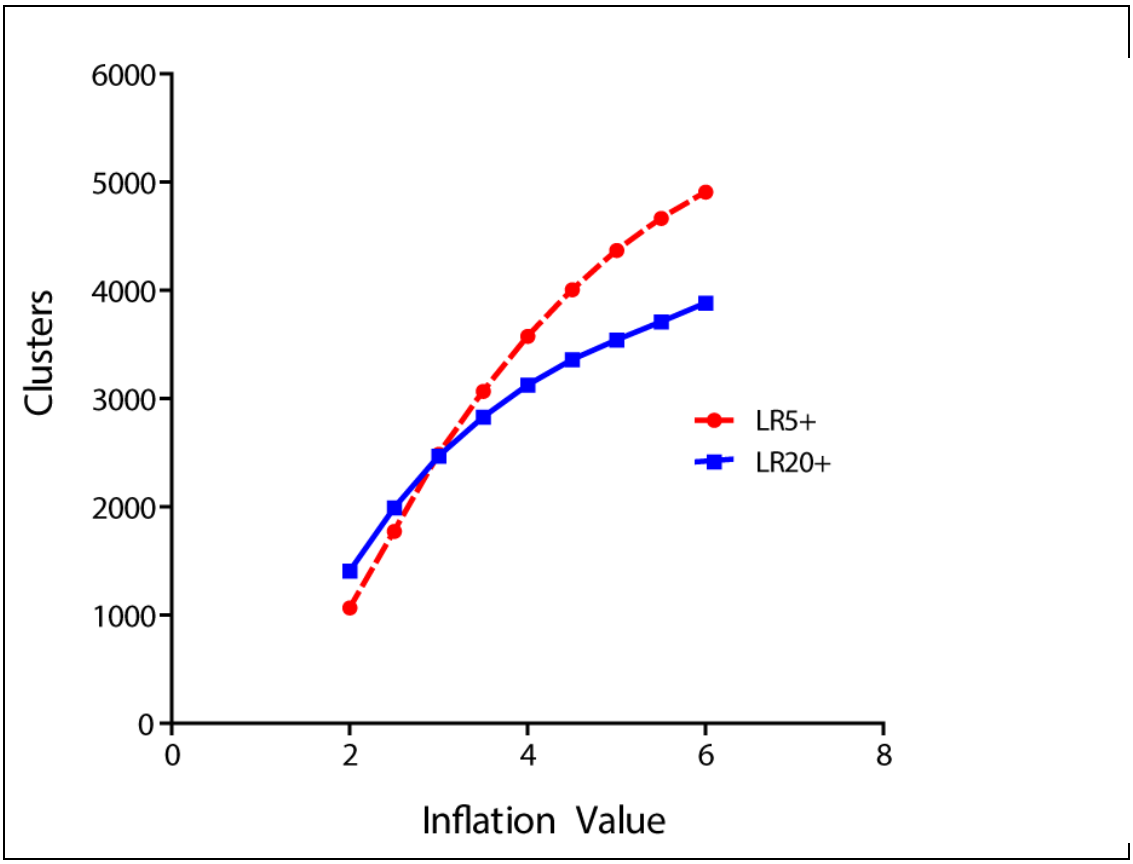

Figure 6-1 Effect of MCL Inflation Value on Cluster Number. Increasing inflation value increases the number of clusters generated but each cluster is correspondingly smaller in size.

Table 6-2 Number of Mapped Clusters in the LR5+ network. The number of mapped genes is provided for each experiment and for each differentially expressed list in parenthesis. The number of clusters identified with each strategy is displayed across inflation values, for each differentially expressed list.

Number of Clusters Mapped with an FDR-corrected p<0.1							
Differentially Expressed List (TvO)	Inflation Value						
TMT (2047 unique mapped GeneIDs)	2	2.5	3	3.5	4	4.5	5
1.5 fold up-regulated (150)	9	8	6	9	8	4	7
1.5 fold down-regulated (232)	6	6	7	8	7	8	8
2 fold up-regulated (30)	1	2	2	2	1	2	2
2 fold down-regulated (67)	3	2	2	3	2	0	3
FDR corrected p<0.1 up-regulated (10)	2	3	2	3	3	3	3
FDR corrected p<0.1 down-regulated (5)	2	1	0	0	0	0	0
FDR corrected p<0.05 up-regulated (9)	2	3	2	3	3	3	3
FDR corrected p < 0.05 down-regulated (4)	3	3	2	1	2	2	1
PAcIFIC (1451 unique mapped GeneIDs)	2	2.5	3	3.5	4	4.5	5
1.5 fold up-regulated (645)	2	1	0	0	0	1	0
1.5 fold down-regulated (539)	2	1	0	0	0	0	0
2 fold up-regulated (543)	33	2	0	1	1	1	1
2 fold down-regulated (405)	4	2	1	0	0	0	0
FDR corrected p<0.1 up-regulated (47)	0	0	0	1	0	1	0
FDR corrected p<0.1 down-regulated (31)	2	2	2	2	2	2	2
FDR corrected p<0.05 up-regulated (47)	0	0	0	1	0	1	0
FDR corrected p < 0.05 down-regulated (27)	1	1	1	1	1	1	1
Number of Clusters Mapped with Cluster >20 Genes, >10% Mapped Genes							
Differentially Expressed List (TvO)	Inflation Value						
TMT (2047 unique mapped GeneIDs)	2	2.5	3	3.5	4	4.5	5
1.5 fold up-regulated (150)	5	3	3	3	3	3	3
1.5 fold down-regulated (232)	5	5	4	3	3	2	3
2 fold up-regulated (30)	0	0	0	0	0	0	0
2 fold down-regulated (67)	1	1	1	1	1	0	2
FDR corrected p<0.1 up-regulated (10)	0	0	0	0	0	0	0
FDR corrected p<0.1 down-regulated (5)	0	0	0	0	0	0	0
FDR corrected p<0.05 up-regulated (9)	0	0	0	0	0	0	0
FDR corrected p < 0.05 down-regulated (4)	0	0	0	0	0	0	0
PAcIFIC (1451 unique mapped GeneIDs)	2	2.5	3	3.5	4	4.5	5
1.5 fold up-regulated (645)	2	1	0	0	0	1	0
1.5 fold down-regulated (539)	2	1	0	0	0	0	0
2 fold up-regulated (543)	8	2	0	1	1	1	1
2 fold down-regulated (405)	2	1	1	0	0	0	0
FDR corrected p<0.1 up-regulated (47)	0	0	0	0	0	1	0
FDR corrected p<0.1 down-regulated (31)	0	0	0	0	0	0	0
FDR corrected p<0.05 up-regulated (47)	0	0	0	0	0	0	0
FDR corrected p < 0.05 down-regulated (27)	0	0	0	0	0	0	0

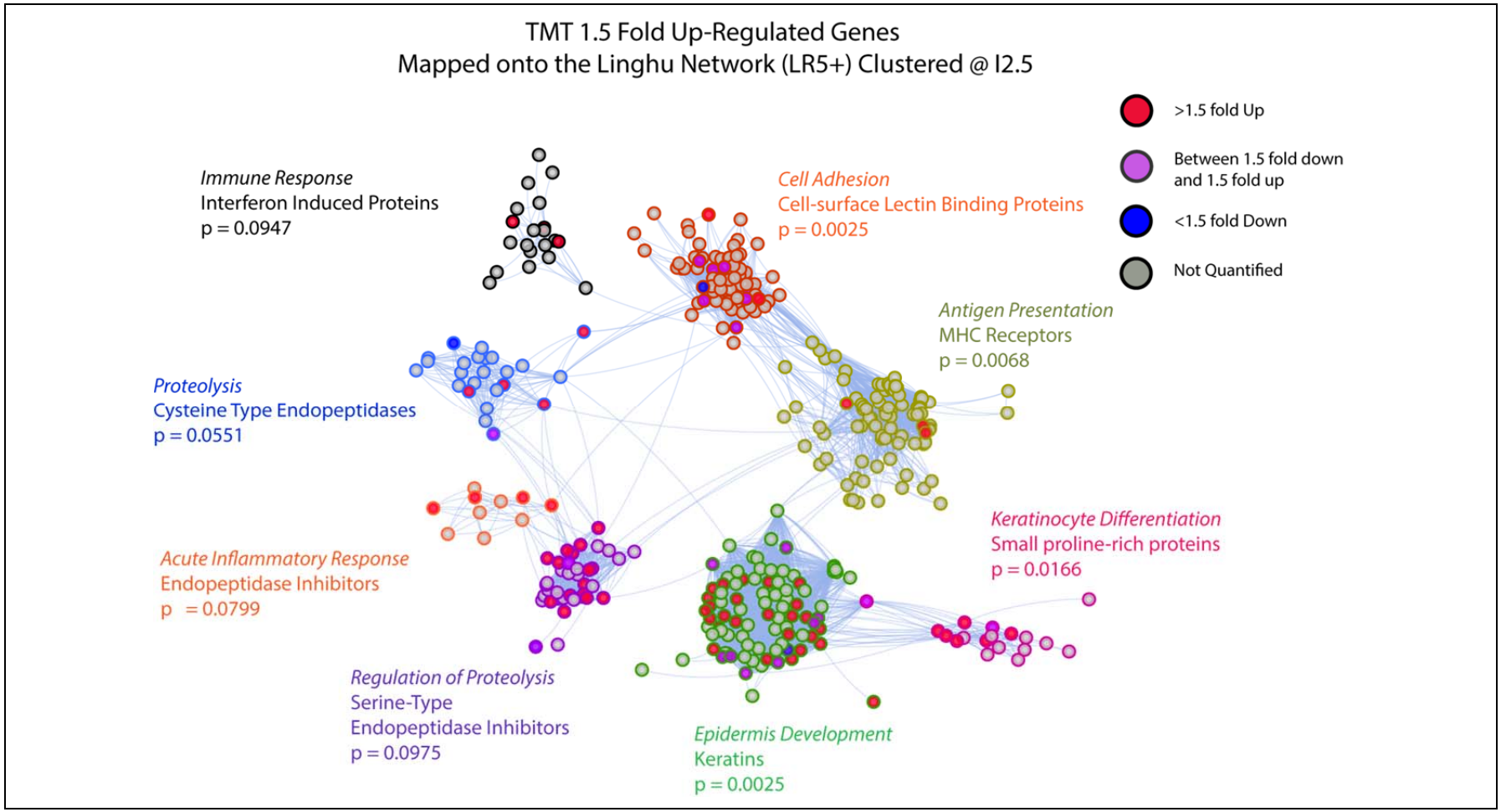

Figure 6-2 Significant Functional Clusters Identified using the TMT1.5 fold Up-Regulated Genes and the Linghu Network. Dysregulated genes were mapped to the LR5+ network clustered at an I-value of 2.5. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original LR5+ network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3).

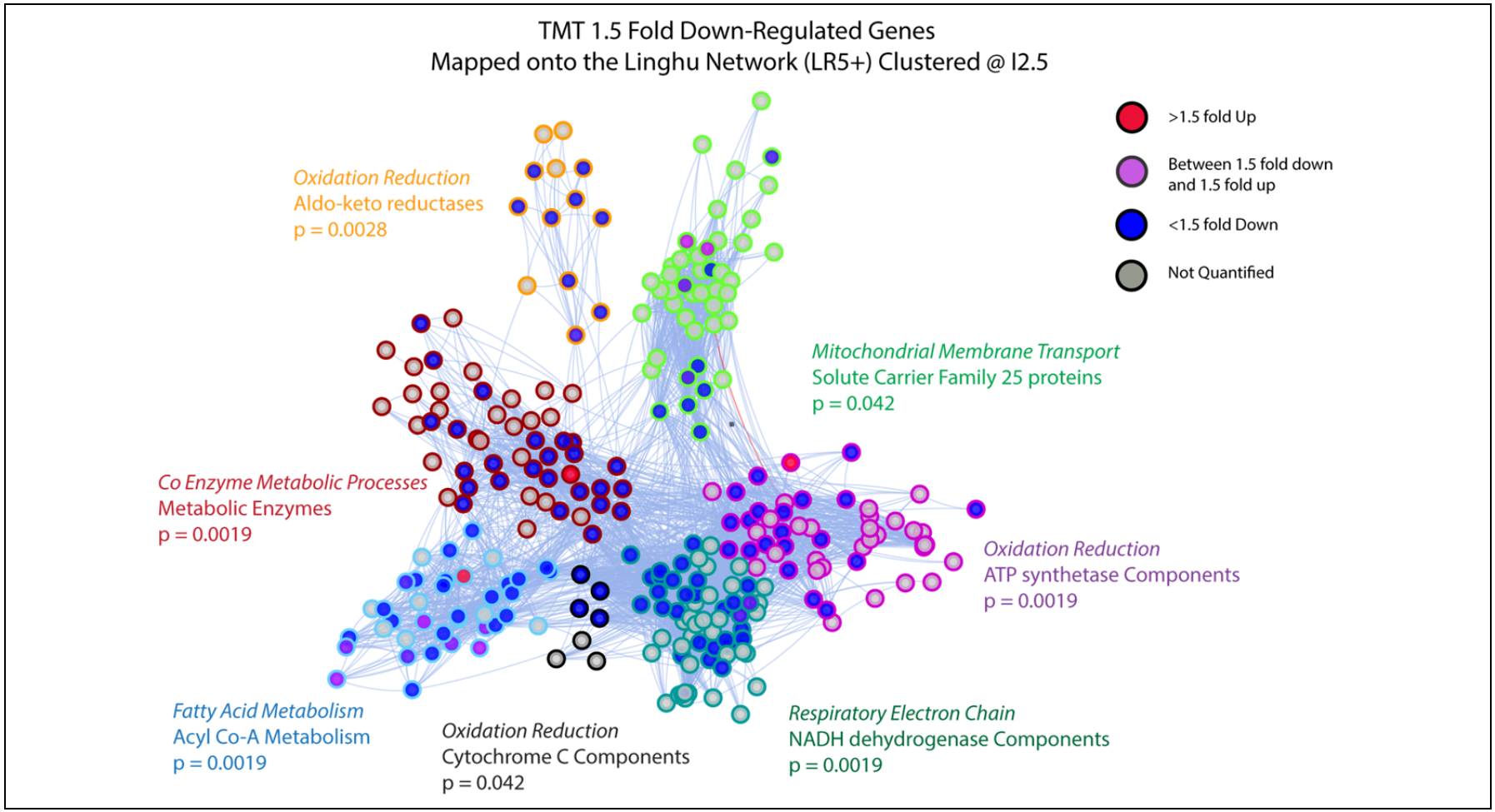

Figure 6-3 Significant Functional Clusters Identified using the TMT1.5 fold Down-Regulated Genes and the Linghu Network. Dysregulated genes were mapped to the LR5+ network clustered at an I-value of 2.5. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original LR5+ network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3).

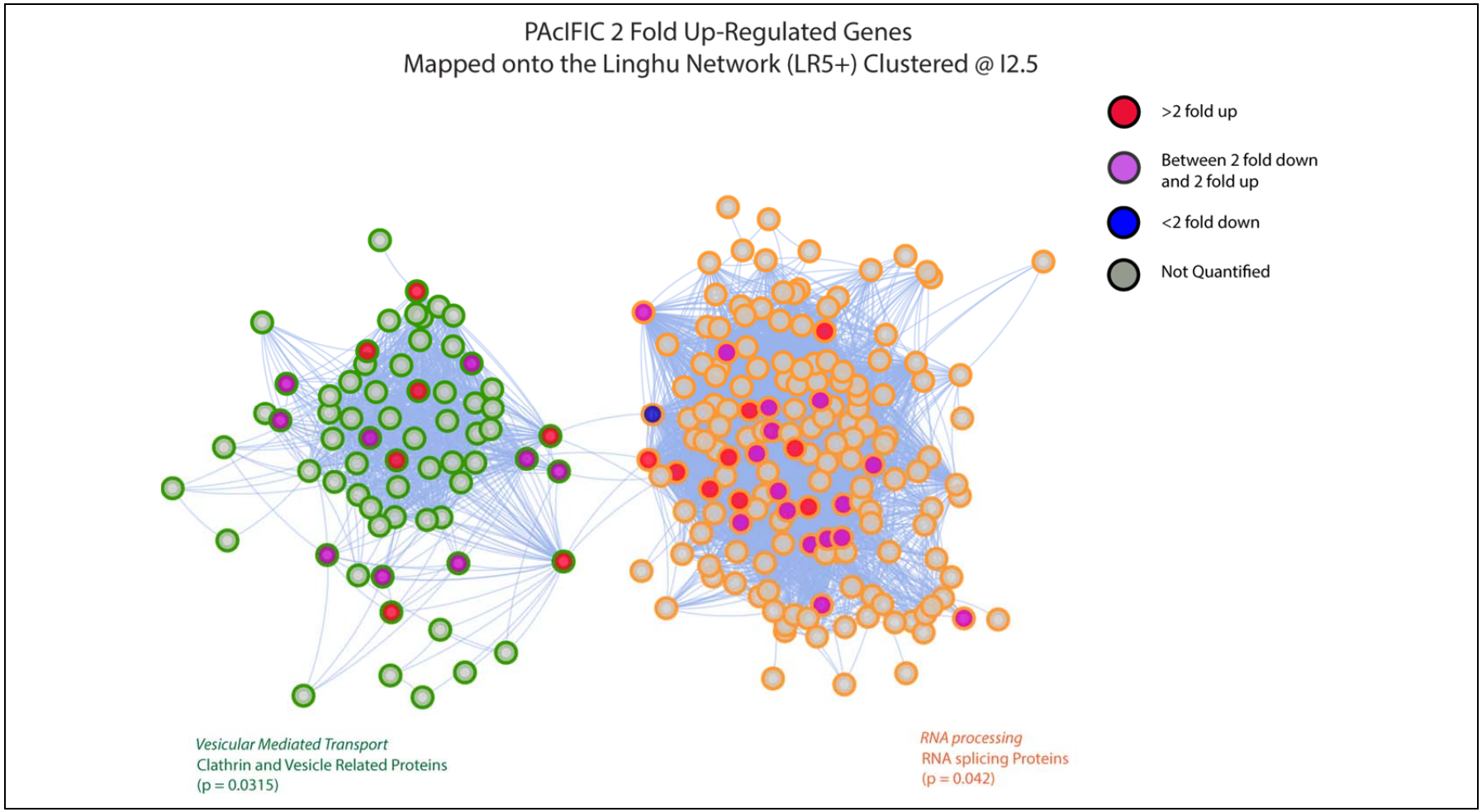

Figure 6-4 Significant Functional Clusters Identified using the PAcIFIC 2 fold Up-Regulated Genes and the Linghu Network. Dysregulated genes were mapped to the LR5+ network clustered at an I-value of 2.5. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original LR5+ network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3).

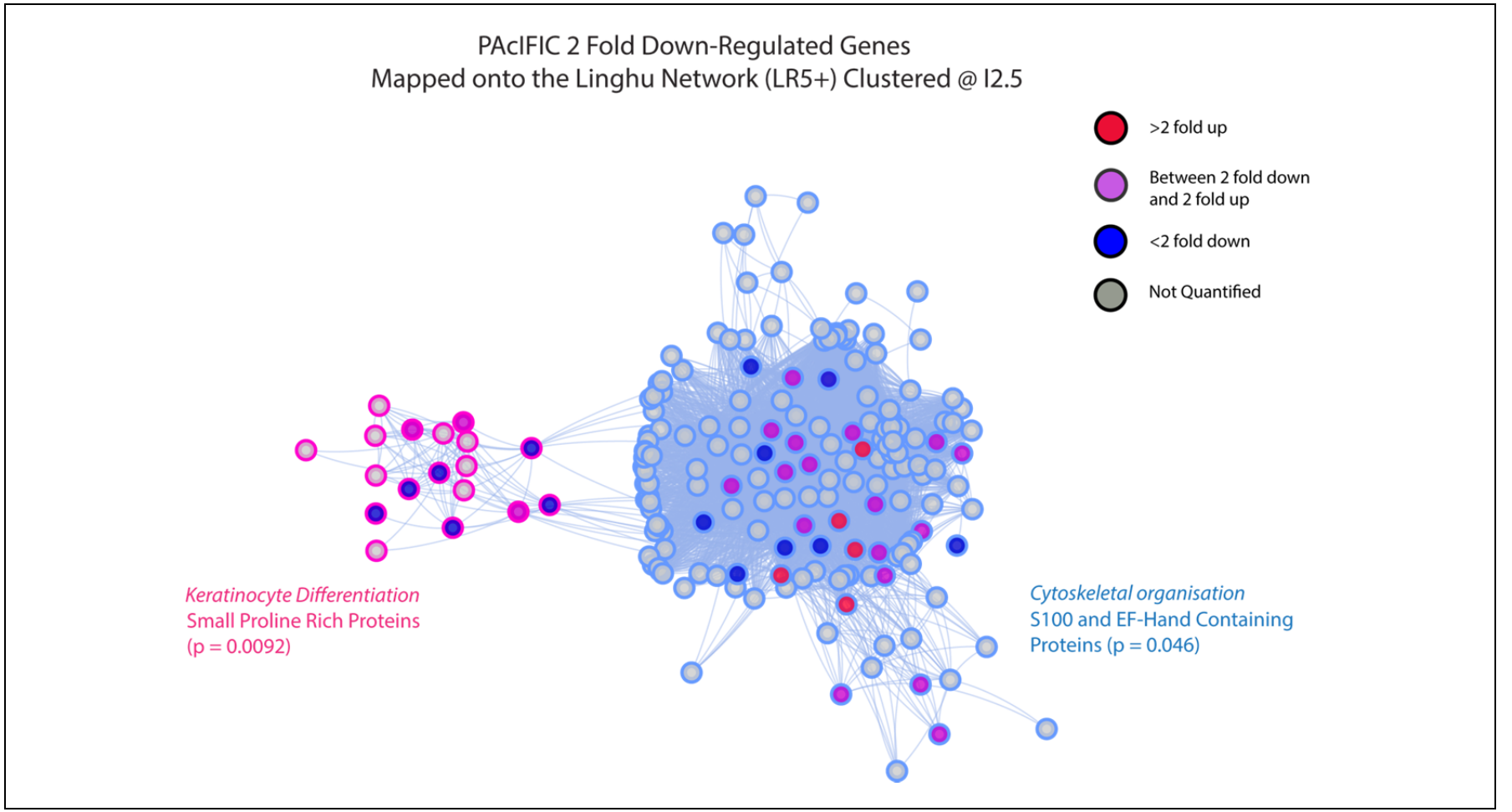

Figure 6-5 Significant Functional Clusters Identified using the PAcIFIC 2 fold Down-Regulated Genes and the Linghu Network. Dysregulated genes were mapped to the LR5+ network clustered at an I-value of 2.5. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original LR5+ network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3).

Table 6-3 Analysis of Significantly Enriched Clusters in the TMT Dataset

TMT 1.5-fold up-regulated genes mapped onto the Linghu network (LR5), clustered @ I2.5.						
Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Up to Top 3 Biological Processes	GO enrichment p-values (FDR-corrected)	Summary Gene Class
23	4	-	95	[cell adhesion] [endocytosis] [immune response]	[1e-17] [6e-10] [2e-5]	Cell surface lectin-binding proteins
32	8	4	76	[immune response] [antigen presentation] [NK-cell activation]	[9e-70] [8e-52] [1e-5]	MHC receptors
129	3	-	20	[immune response] [response to virus]	[5e-5] [6e-4]	Interferon induced proteins
108	4	2	24	[proteolysis]	[1e-20]	Cysteine-type Endopeptidases
70	17	1	38	[regulation of proteolysis] [wound healing] [coagulation]	[2e-7] [6e-6] [2e-4]	Serine-type Endopeptidase inhibitors
302	4	-	10	[acute inflammatory response]	[ns]	Endopeptidase Inhibitors
25	33	3	90	[intermediate filament-based process] [ectoderm development] [epidermis development]	[3e-24] [2e-22] [1e-16]	Keratins
141	8	-	18	[keratinocyte differentiation] [epidermal cell differentiation] [keratinization]	[3e-35] [2e-34] [4e-31]	Small-proline-rich proteins
TMT 1.5-fold down-regulated genes mapped onto the Linghu network (LR5), clustered @ I2.5.						
Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Up to Top 3 Biological Processes	GO enrichment p-values (FDR corrected)	Summary Gene Class
51	2	25	51	[coenzyme metabolic process] [energy generation] [acetyl-CoA metabolic process]	[2e-24] [2e-22] [2e-20]	Metabolic enzymes
52	1	20	51	[proton transport] [ATP synthesis] [ion transmembrane transport]	[1e-45] [3e-42] [4e-40]	ATP synthetase components
58	1	10	44	[transmembrane transport] [mitochondrial transport]	[8e-60] [1e-41]	Solute carrier family 25 proteins
69	1	27	38	[fatty acid metabolism] [carboxylic acid catabolism]	[1e-24] [8e-19]	Acyl-CoA metabolism
226	-	9	13	[oxidation reduction] [cellular aldehyde metabolism]	[2e-14] [1e-5]	Aldo-keto reductases
412	-	4	7	[oxidation reduction]	[4e-5]	Cytochrome-c components
39	-	30	62	[mitochondrial electron transport] [respiratory electron chain] [ATP synthesis]	[3e-102] [1e-100] [2e-96]	NADH dehydrogenase components

Table 6-4 Analysis of Significantly Enriched Clusters in the PAcIFIC Dataset

PAcIFIC 2 fold up-regulated genes mapped onto the Linghu network (LR5), clustered at I2.5.						
Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Up to Top 3 Biological Processes	GO enrichment p-values (FDR corrected)	Summary Gene Class
11	21	5	151	[RNA splicing] [mRNA processing] [RNA splicing, via trans-esterification]	[9e-22] [8e-20] [4e-11]	RNA splicing
35	15	1	63	[vesicle mediated transport] [intracellular protein transport] [protein localisation]	[3e-70] [2e-62] [2e-60]	Clathrin and vesicle related
PAcIFIC 2 fold down-regulated genes mapped onto the Linghu network (LR5), clustered at I2.5.						
Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Up to Top 3 Biological Processes	GO enrichment p-values (FDR corrected)	Summary Gene Class
141	-	9	17	[keratinocyte differentiation] [epidermal cell differentiation] [keratinocyte differentiation]	[3e-35] [1e-34] [4e-31]	Small proline rich proteins
8	12	19	176	[cytoskeletal organisation] [calcium ion transport] [detection of stimuli]	[3e-9] [1e-6] [4e-6]	S100 and EF-hand domain containing proteins

6.3.4 Evaluation of the Stein Network

To the author's knowledge, the Linghu network represents the highest genome coverage functional linkage network available. To verify the findings from the Linghu FLN, an independent published FLN was used; the Stein network[670]. This high coverage human functional linkage network was constructed using human[651,652,671], fly[652], worm[652] and yeast[652] protein- protein interaction databases, domain-domain interations[672], gene co-expression[661], shared gene ontology biological process annotations[673], and text mining for published protein-protein interactions not present in other databases[674]. The complete and filtered Linghu networks and the Stein network are compared in Table 6-5.

Table 6-5 Comparison of Human Functional Linkage Networks

Network	Network Node Identifier	Nodes	Edges	Edge Represents
Linghu (LR 1+)	GeneID	21,657	22,388,609	Functional Linkage (16 domains)
Linghu (LR5+)	GeneID	16,231	530,388	Functional Linkage (16 domains)
Linghu (LR20+)	GeneID	10,342	69,649	Functional Linkage (16 domains)
Stein (Complete)	Uniprot AC	10,957	209,988	Functional Linkage (9 domains)

All quantified and dysregulated proteins from both mFASP TMT and PAcIFIC experiments were mapped against the unique Uniprot Accessions (AC) in the Stein network (Table 6-6). The MCL method was again used to cluster the Stein network at I-values 2.0 to 6.0. Significant clusters were identified for the TMT 1.5 fold and PAcIFIC 2 fold dysregulated gene lists in the same fashion as for the Linghu network (Table 6-7). For the PAcIFIC data, an I-value of 2.5 and for the TMT data, an I-value of 4 were selected as optimal. The cluster members and cluster edges were extracted from the Stein network and were used to create network visualisations in Cytoscape, in the same manner as the enriched LR5+ clusters from the Linghu network (Figure 6-6 to Figure 6-8). The biological process ontologies for the significantly enriched clusters are summarised in Table 6-8 and Table 6-9.

Table 6-6 Mapping of TMT and PAcIFIC Datasets to the Stein Network

	All TMT	All PAcIFIC	TMT 1.5 fold up	TMT 1.5 fold down	PAcIFIC 2 fold up	PAcIFIC 2 fold down
Mapping Protein List	2,058	1,472	154	232	543	408
Stein Protein List	10,957	10,957	10,957	10,957	10,957	10,957
Mapped Proteins (%)	1,565 (76%)	980 (67%)	114 (74%)	162 (69%)	348 (64%)	278 (68%)

Table 6-7 Enriched Clusters in the Stein Network
For each differentially expressed list, the number of Uniprot ACs searched is provided in parenthesis.

Significant Enriched Clusters (FDR-corrected p<0.1)									
Differentially Expressed List (TvO)	Inflation Value								
TMT (1565 unique mapped Uniprot ACs)	2	2.5	3	3.5	4	4.5	5	5.5	6
1.5 fold up-regulated (114)	5	5	2	3	6	6	6	6	5
1.5 fold down-regulated (162)	8	9	7	8	7	7	8	7	7
PAcIFIC (980 unique mapped Uniprot ACs)	2	2.5	3	3.5	4	4.5	5	5.5	6
2 fold up-regulated (348)	1	1	0	1	0	0	0	0	0
2 fold down-regulated (278)	1	1	1	0	0	0	1	0	0

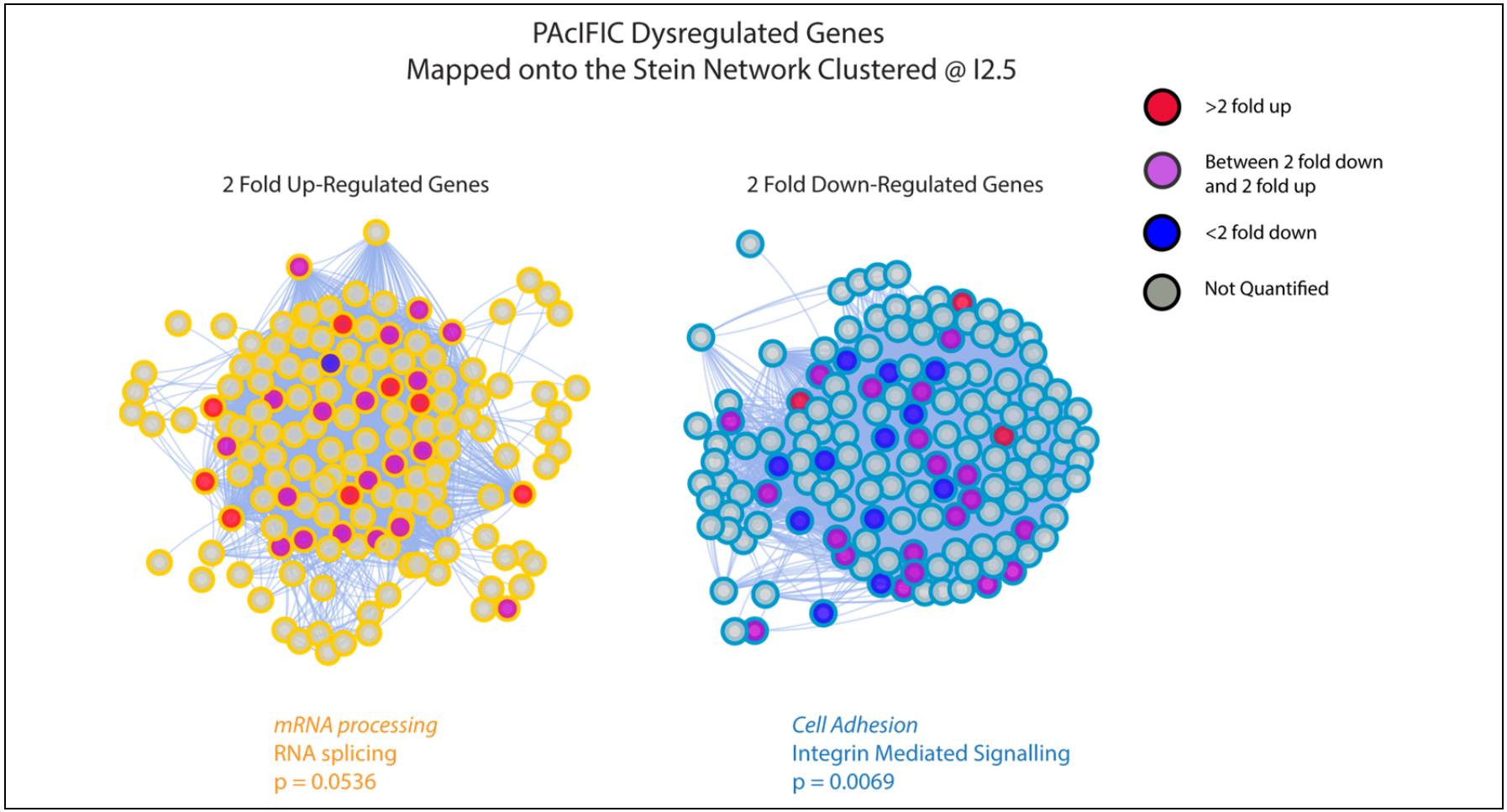

Figure 6-6 Significantly Enriched Clusters Identified in the PAcIFIC Dataset Mapped onto the Stein Network. Gene lists were separately mapped to the Stein network clustered at an I-value of 2.5. Enriched clusters are displayed together for comparison. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original Stein network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3)

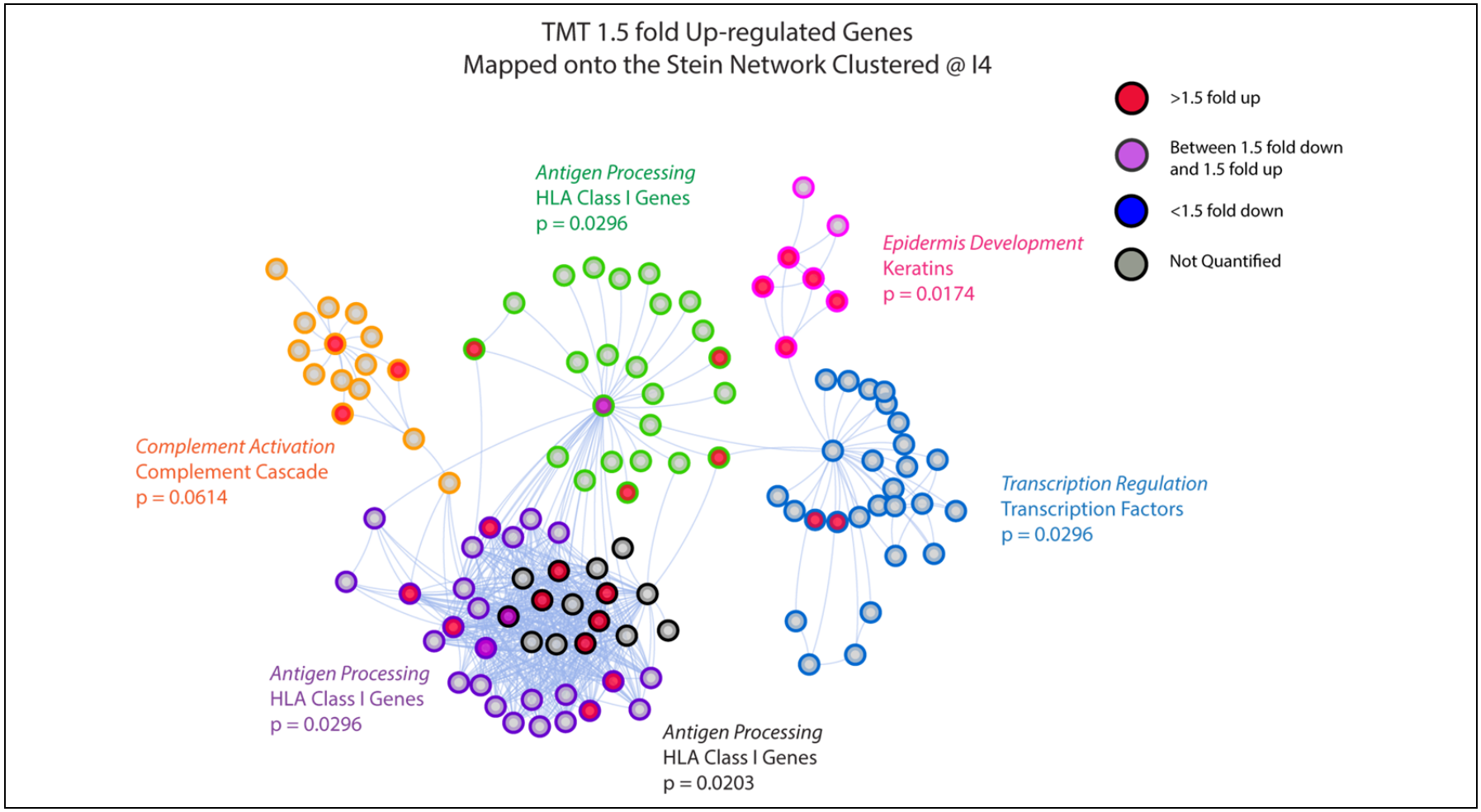

Figure 6-7 Significantly Enriched Clusters Identified in the TMT 1.5 fold Up-Regulated Genes Mapped to the Stein Network. Genes were mapped to the Stein network clustered at an I-value of 4.0. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original Stein network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3)

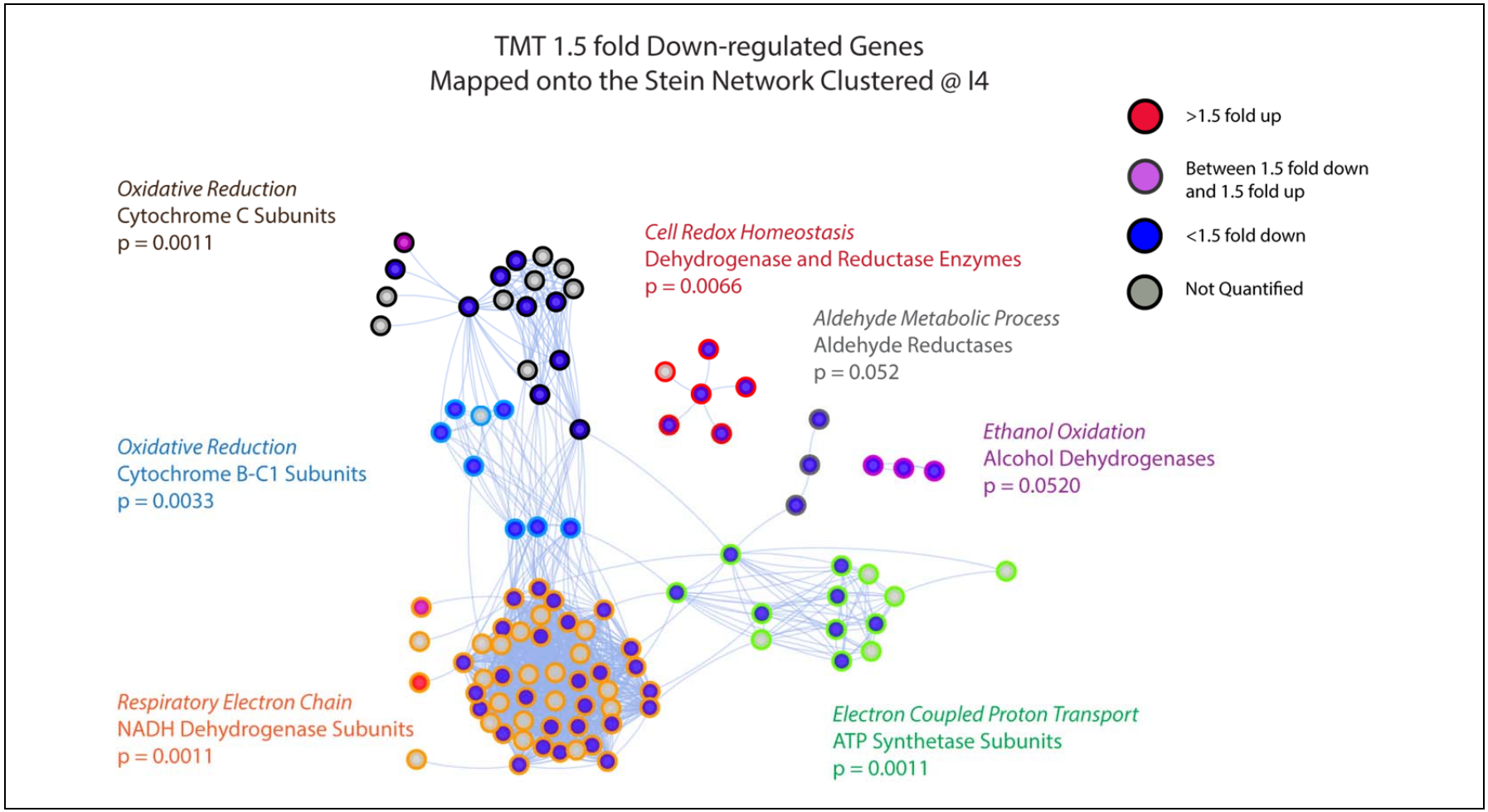

Figure 6-8 Significantly Enriched Clusters in the TMT 1.5 fold Down-Regulated Genes Mapped to the Stein Network. Genes were mapped to the Stein network clustered at an I-value of 4.0. Nodes have been filled according to the TvO expression value (see legend). Node outline corresponds to the cluster membership from the original Stein network. The biological process ontology (italics), a summary description of the cluster genes, and the FDR-corrected p-value for the cluster enrichment are provided for each cluster (see Methods 2.13.2 and 2.13.3)

Table 6-8 Enriched Clusters in the Stein Network Identified using the TMT Dataset

TMT 1.5 fold up-regulated genes mapped onto the Stein network, clustered @ I4.						
Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment p-values (FDR corrected)	Summary Gene Class
276	5	-	7	[epidermis development] [ectoderm development] [intermediate filament-based process]	[2e-7] [1e-7] [7e-4]	Keratins
42	2	-	27	[transcription regulation] [regulation of RNA metabolic process]	[5e-6] [3e-6]	Transcription factors
55	4	1	24	[antigen processing and presentation of peptide antigen via MHC class I] [antigen processing and presentation] [immune response]	[5e-5] [3e-4] [7e-3]	HLA Class I genes
83	6	-	15	[antigen processing and presentation via MHC class I]	NA (proteins all from one gene)	HLA Class I genes
37	6	-	25	[antigen processing and presentation of peptide antigen via MHC class I] [antigen processing and presentation] [immune response]	[4e-11] [3e-11] [4e-7]	HLA Class I genes
99	3	-	15	[complement activation] [humoral response] [activation of immune response]	[2e-17] [1e-15] [4e-15]	Complement cascade
TMT 1.5 fold down-regulated genes mapped onto the Stein network, clustered @ I4.						
Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment p-values (FDR corrected)	Summary Gene Class
19	-	30	48	[mitochondrial electron transport] [ATP synthesis coupled electron transport] [electron transport chain]	[1e-101] [8e-94] [1e-89]	NADH dehydrogenases
74	-	10	74	[generation of precursor metabolites and energy] [cellular respiration] [oxidative reduction]	[3e-10]	Cytochrome-c subunits
120	-	8	13	[proton transport] [energy coupled proton transport] [ATP biosynthetic transport]	[4e-25] [4e-24] [1e-23]	ATP synthetase subunits
219	-	7	8	[electron transport chain] [mitochondrial electron transport] [generation of precursor metabolites and energy]	[8e-14] [4e-13] [3e-12]	Cytochrome b-c1 subunits
322	-	5	6	[cell redox homeostasis] [oxidation reduction] [coenzyme metabolic process]	[8e-5] [9e-4] [2e-2]	Dehydrogenase and reductase enzymes
918	-	3	3	[ethanol oxidation] [drug metabolic process]	ns	Alcohol dehydrogenases
929	-	3	3	[cellular aldehyde metabolic process]	ns	Aflatoxin aldehyde reductases

Table 6-9 Enriched Clusters in the Stein Network Identified using the PAcIFIC Dataset

Cluster Ref No.	Genes TvO >1	Genes TvO <1	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment p-values (FDR corrected)	Summary Gene Class
PAcIFIC 2 fold up-regulated genes mapped onto the Stein network, clustered @ I2.5.						
5	21	7	137	[RNA splicing] [mRNA processing] [spliceosome assembly]	[5e-207] [2e-187] [2e-27]	RNA splicing
PAcIFIC 2 fold down-regulated genes mapped onto the Stein network, clustered @ I2.5.						
7	9	26	114	[cell adhesion] [integrin-mediated signalling pathway] [extracellular matrix organisation]	[7e-84] [4e-42] [9e-28]	Collagens and integrins

The enriched clusters from the Stein network showed highly similar biological functions to the enriched clusters from the Linghu LR5+ network. These networks, although similar in scope, used different sources and methods to define the network edges. Therefore the clusters generated by MCL would be expected to contain different nodes. Identifying clusters characterised by similar biological process ontologies provides strong support for the dysregulation of these pathways in OAC compared to normal squamous tissue.

6.4 Discussion

This study has adapted strategies originally developed for gene-set and gene-network enrichment analysis to identify functional pathways from incomplete proteomic datasets. Genome-scale functional linkage networks[460,670] were clustered to generate biological functional clusters – verified by the significant enrichment of DAVID biological process ontologies among the members of each cluster. After iteration of the optimal network size and cluster inflation-value, significantly enriched clusters could be identified from both mFASP TMT and PAcIFIC datasets. In the TMT up-regulated gene list, the identified clusters were enriched for antigen processing and immune regulatory genes. The original samples were biopsies of human tumours, and therefore proteins may be detected from epithelial-derived tumour cells, infiltrative immune cells or other stromal components.

Tumour associated inflammation has been demonstrated in many cancers and in some cases may drive the oncogenic phenotype[675,676]. The complexities of the immune – cancer cell synapse are still being uncovered with evidence for a symbiotic relationship with tumour cell mediated recruitment of bone marrow derived immune cells[677] and tumour-associated macrophages promoting the oncogenic phenotype of epithelial derived tumour cells[675,678,679]. The observed enriched processes may therefore reflect an upregulation of genes in stromal immune cells or in the epithelial-derived cells.

A transgenic mouse model with constitutively active IL-1β expressed in the oesophagus and squamous fore-stomach phenocopies the human Barrett's carcinogenesis sequence[102]. This suggests oesophageal inflammation is critical in the development of OAC. Modulating this axis may lead to successful preventative and therapeutic strategies and exploitation of the tumour-immune cell synapse as a druggable interface is an emerging field.

Intriguingly a large cluster of metabolic genes associated with aerobic metabolism were identified from the down-regulated gene list. It could be speculated this is a consequence of the Warburg effect[680]. Greater numbers of samples would be required to confirm this as a generalised phenomenon.

Fewer significant clusters could be identified from the PAcIFIC dataset but similar to the TMT dataset, significant changes in keratin-related genes were observed between cancer and normal squamous tissues. The observation of significantly up-regulated RNA processing and splicing-related genes may provide support for the observation of alternative spliced and novel transcripts between cancer and normal tissues[681,682].

The enriched clusters derived from the Linghu network were interpreted as representing enriched biological processes in the dysregulated gene lists. The cluster membership will be a direct consequence of the original network topology and therefore the identified biological processes may change with network connectivity. An independent human functional linkage network, the Stein network, analysed in the same manner identified highly similar biological processes[670]. This adds strength to the conclusion that these processes are truly dysregulated in OAC tissue. A caveat for this approach is that well studied genes and their interactors are over-represented in the databases used to construct both networks. This leads to higher network connectivity for these genes and a greater likelihood of representation in the clusters. Novel biological processes in contrast will not be identified and would require alternative approaches.

For both the Linghu and Stein FLNs, the TMT data provided the greatest number of enriched clusters and number of genes mapped to each cluster. To evaluate if these findings and other novel functional clusters could be observed in a larger cohort, a further 5 patient's samples were processed by mFASP and then subjected to TMT LC-MS/MS using the same method as the two patient's samples in this analysis. The analysis of the combined mFASP TMT data will be considered in the final chapter.

Chapter 7: Application of the Optimised Proteomic Method to OAC and Matched Normal Tissues

7.1 Chapter Aims

a) To combine the results from seven patients' samples analysed by the optimised mFASP TMT method.

b) To define the technical and biological variability of this method.

c) To define dysregulated and potentially tissue-specific proteins using expression data from OAC, normal oesophagus and normal stomach.

d) To undertake a network analysis of these quantitative data to validate previous findings and uncover targetable pathways.

7.2 Introduction

Previous chapters have confirmed accurate, sensitive quantitative protein expression data can be produced from human oesophageal tissue samples by shotgun proteomics. The optimised mFASP TMT method was associated with the greatest proteome coverage and quantitative dynamic range. A network analysis using data from two patients identified significantly dysregulated biological processes. To validate and extend these findings, data was generated from a further five patients' samples using the same sample preparation and proteomic analysis method. This chapter will consider the data for all seven patients' tissues processed by mFASP TMT.

Differential expression is commonly considered between two tissue types. The comparisons between adenocarcinoma and normal squamous epithelium in the previous studies revealed many dysregulated proteins, some of which represent proteins associated with glandular differentiation and some associated with oncogenesis. Glandular associated proteins may be expressed in gastric and intestinal epithelium and may not represent tractable targets for therapy as toxicity due to intestinal epithelial damage would be expected.

To enhance the probability of identifying tissue-specific proteins, patient-matched normal gastric tissues were included with OAC and normal squamous tissues. Multi-tissue proteomic profiling has recently been described using mouse tissue and relative quantitation by a super-SILAC approach[683]. In this study, snap-frozen biopsies from 28 tissue types were subjected to shotgun proteomics with a spike-in, heavy-labelled mixture of all tissues obtained from the SILAC mouse. By comparing the relative expression of proteins across tissues, tissue-specific expression could be highlighted. The oesophagus was not included in this profiling effort although gastrointestinal tissues with columnar epithelia showed similar expression patterns. This suggests that including columnar epithelium-lined gastric tissue along with squamous and OAC tissue may allow the separation of common "glandular" associated proteins from true cancer-associated proteins.

7.3 Results

7.3.1 TMT reporter ion intensities

The clinical data for each patient donating tissue for this study are summarised in Table 7-1. Peptides from each tissue type from a single patient had been labelled independently with two TMT reporters (Figure 5-2). These technical replicates were analysed independently.

Table 7-1 Clinical Characteristics of Patients Donating Tissue for mFASP TMT proteomics

	Patient (P)						
	P44	P46	P48	P51	P53	P60	P61
Protocol	mFASP TMT	mFASP TMT	mFASP TMT	mFASP TMT	mFASP TMT	mFASP TMT	mFASP TMT
Gender	Male	Male	Male	Male	Female	Male	Male
Age	59	67	65	41	52	60	58
Histology	ACC	ACC	ACC	ACC	ACC	ACC	ACC
Location	OGJ Type II	OGJ Type I	OGJ Type II	OGJ Type I	OGJ Type I	OGJ Type II	Oes Lower
Neoadjuvant Therapy	2xCF	2xCF	2xCF	2xCF	2xCF	2xCF	2xCF
Post-Chemotherapy CT	Stable Disease	Stable Disease	Stable Disease	Stable Disease	Stable Disease	Stable Disease	Stable Disease
Surgery	ILO	ILO	ILO	ILO	ILO	ILO	ILO
Maximum Tumour Diameter	38 mm	70 mm	40 mm	50 mm	83 mm	52 mm	35 mm
Distance to PRM and DRM	>1 mm	>1 mm	>1 mm	>1 mm	>1 mm	>1 mm	>1 mm
Distance to CRM	4.2 mm	0.0 mm	0.0 mm	0.3 mm	3.0 mm	1.0 mm	1.0 mm
Differentiation	Moderate	Poor	Poor	Poor	Poor	Poor	Poor
Lymphovascular Invasion	Y	Y	Y	Y	N	Y	Y
Venous Invasion	N	N	N	Y	N	N	N
Perineural Invasion	N	Y	Y	Y	N	Y	Y
T stage	T2	T4a	T3	T3	T2	T3	T3
Number of Positive Nodes	2	8	16	7	1	7	3
Total Number Nodes Resected	27	18	28	28	23	21	37
AJCC Stage	IIB	IIIC	IIIC	IIIC	IIB	IIIC	IIIB
Alive at Time of Analysis	Yes	No	No	Yes	Yes	No	Yes
Overall Survival From Diagnosis	18.2 months (censored)	15.3 months	10.5 months	13.9 months (censored)	17.8 months (censored)	17.8 months	14.5 months (censored)

Abbreviations: 2xCF - 2 cycles of Cisplatin and 5-Fluorouracil, ILO – Ivor-Lewis Oesophagectomy, AJCC – American Joint Committee on Cancer, OGJ – Oesophagogastric Junctional Tumour, PRM – Proximal resection margin, DRM – Distal resection margin, CRM – Circumferential resection margin., Y-Yes, N-No, CT – Computed Tomography.

The TMT reporter ion fold changes of over 10^7 were detected confirming the high quantitative dynamic range of this method (Table 7-2). An example distribution of tumour peptide reporter ion intensities is shown in Figure 7-1. This shows a complex bimodal distribution for the raw reporter intensities. This was apparent for all the tissue types analysed. This may reflect an enrichment of low abundance peptides produced by OFF-GEL fractionation. The ratio of reporters, however, followed a Gaussian distribution (C) and therefore parametric statistics were used to assess the ratios between tissue types.

Table 7-2 Dynamic Range of TMT Reporter Intensities (Peptide FDR 1%, 1 Unique Peptide per Protein)

Patient Samples	P44	P46	48	P51	P53	P60	P61
Maximum Reporter Ion Fold Change	$7.7x10^7$	$7.7x10^7$	$2.12x10^7$	$9.09x10^7$	$3.44x10^7$	$5.87x10^7$	$2.38x10^7$

7.3.2 Assessment of Technical Variation

The distribution of technical replicate ratios for each experiment is summarised in Table 7-3. This demonstrates the small degree of systematic error in the ratios from most experiments. This error is likely to be secondary to slight variations in the total amount of peptide from each sample labelled with each TMT reporter. After median normalisation, the $\log_2$ ratios all followed an approximately Gaussian distribution centred on 0 (Figure 7-2).

The variability in technical replicate ratios from each experiment and the combination of all experiments is shown in Figure 7-3. The combination of technical replicate ratios from all experiments provides a robust estimate of the technical variability and provides a possible method to determine a statistical threshold signifying differential expression. Using this method, a median normalised $\log_2$ expression ratio greater than 0.8515 or less than -0.8515 from the mean would be predicted to have a <5% probability of arising through technical variation alone. The corresponding 1% and 0.1% confidence limits are demonstrated in Table 7-4.

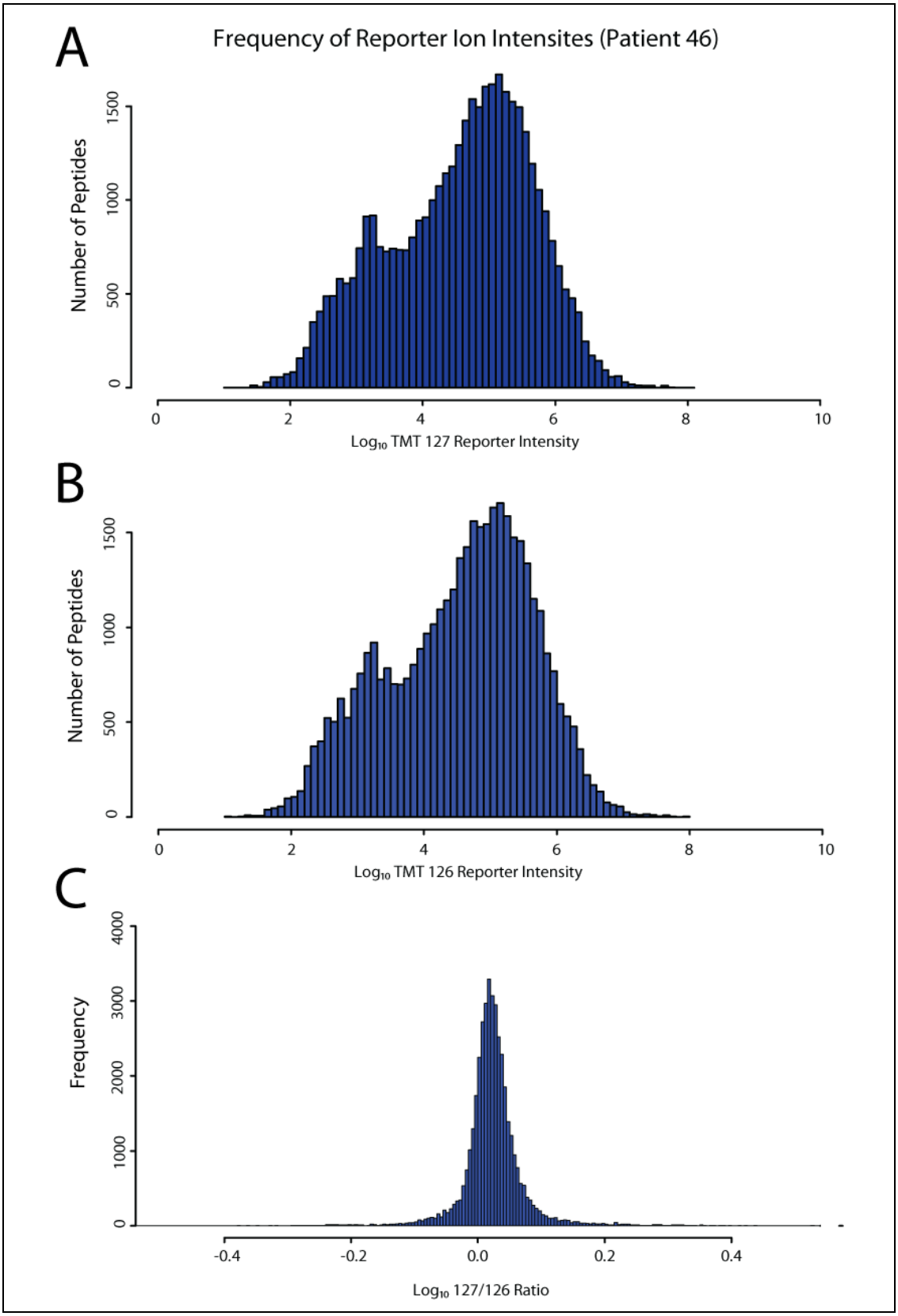

Figure 7-1 Example Distributions of Peptide Reporter Intensities. A – 127 reporter intensity. B - 126 reporter intensity. C – Ratio of 127/126 intensities.

Table 7-3 Technical Replicate Ratios by Experiment. Each patient tissue sample was labelled in technical duplicate. The Mean, Median and standard deviation of raw and Log_2 ratios of peptide reporter ion intensities is reported for each sample.

Reporter Ratio (P = Patient)	Median raw ratio	Mean raw ratio	Standard Deviation (SD)	Mean Log_2 ratios	SD of Log_2 ratios	Number of ratios
P46 126/127	0.9522869	0.9424203	0.1209093	-0.1058829	0.2263472	40872
P46 128/129	0.9212349	0.9871393	0.1063501	-0.1228993	0.1685073	40861
P46 130/131	1.089402	1.157752	0.4982529	0.1182558	0.4093816	40447
P44 126/127	0.9214896	0.9542108	0.2071119	-0.1058829	0.2263472	37862
P44 128/129	1.00383	0.924225	0.1988749	-0.06216656	0.4231987	37737
P44 130/131	1.096927	1.132269	0.4987867	0.1681628	0.2956012	37149
P48 126/127	1.519258	1.472785	0.3015153	0.5190464	0.376901	43892
P48 128/129	0.9099537	0.975992	0.4419535	-0.1146072	0.496436	42480
P48 130/131	1.647579	1.654255	0.6915029	0.6566526	0.474313	42343
P51 126/127	1.021262	1.011923	0.1523257	-0.000425345	0.23806	40838
P51 128/129	1.026764	1.188478	0.6881133	0.1359064	0.517445	39768
P51 130/131	0.7734064	0.8630549	0.6577253	-0.3010062	0.378115	39387
P53 126/127	0.9662025	0.9532796	0.1298938	-0.08014842	0.1944109	40786
P53 128/129	1.304138	1.279188	0.2433199	0.3180678	0.3874888	40337
P53 130/131	1.360151	1.367582	0.3605723	0.4310588	0.2251194	39933
P60 126/127	1.116098	1.102766	0.1609282	0.1235559	0.244673	34014
P60 128/129	0.8904177	0.8918027	0.1081798	-0.1788884	0.225214	34050
P60 130/131	0.7532126	0.8004937	0.4092182	-0.3634508	0.265779	33478
P61 126/127	1.453871	1.431918	0.1848371	0.5027554	0.234084	35377
P61 128/129	0.9727497	0.9595118	0.1423883	-0.08037295	0.2790071	35282
P61 130/131	1.305033	1.345982	0.5784023	0.3787494	0.3411583	34916
All Ratios	**1.023174**	**1.119116**	**0.4491993**	**0.09435871**	**0.4344242**	**811809**

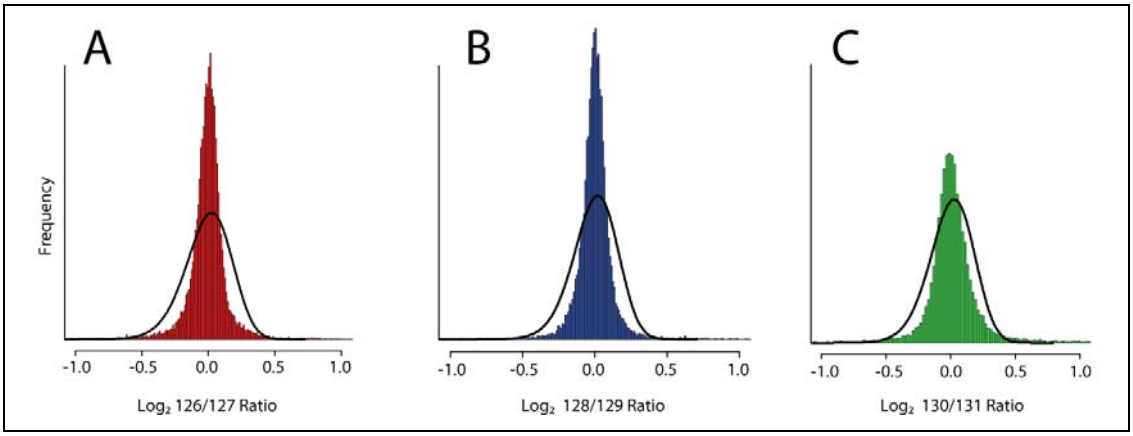

Figure 7-2 Example Distribution of Median Normalised Reporter Ratios from Patient 46. Technical replicate ratios are reported from adenocarcinoma (A), normal oesophagus (B) and normal stomach (C). Theoretical normal distributions with the same mean and SD are displayed as solid black lines highlighting the leptokurtotic distributions.

Figure 7-3 Technical Variation by Experiment. The technical replicate ratios from each tissue from each patient are displayed. Points represent the mean ratio after median normalisation. Bars represent the standard deviation.

Table 7-4 Confidence Intervals for Differential Expression Using the Combined Dataset

Probability of arising through technical variation	Confidence Intervals around the Mean Log_2 Ratio
P<0.05	0.8514558, -0.8514558
P<0.01	2.326348, -2.326348
P<0.001	3.090232, -3.090232

The relatively low degree of technical variability is emphasised when the ratios of technical replicates for single patient's samples are compared (Figure 7-4A). This high correlation of Log_2 ratios between technical replicates was present across all experiments (Figure 7-4B, median Pearson correlation coefficient [PCC] = 0.9811, $p<0.001$). There was also a generally good correlation between biological replicates when TvO ratios were considered (median PCC = 0.555, $p<0.001$) but not between TvO and TvG ratios (median PCC = 0.0115, $p>0.05$). This underscores the diversity, at least in protein expression, of the tissues studied. The good correlation between biological replicates is reassuring and demonstrates concordance of the proteomes from histologically similar tissues.

Unsupervised hierarchical clustering was performed to further examine the differences between biological and technical replicates (Figure 7-5). Technical replicates were all clustered as pairs with a suggestion of a subgroup consisting of patients 46, 61 and 44 although the distance on the dendrogram to the other patients was small. No obvious clinical features were identified in these three cases compared to the other patients (Table 7-1) although this study is underpowered to detect the effect of subtle clinical differences on protein expression.

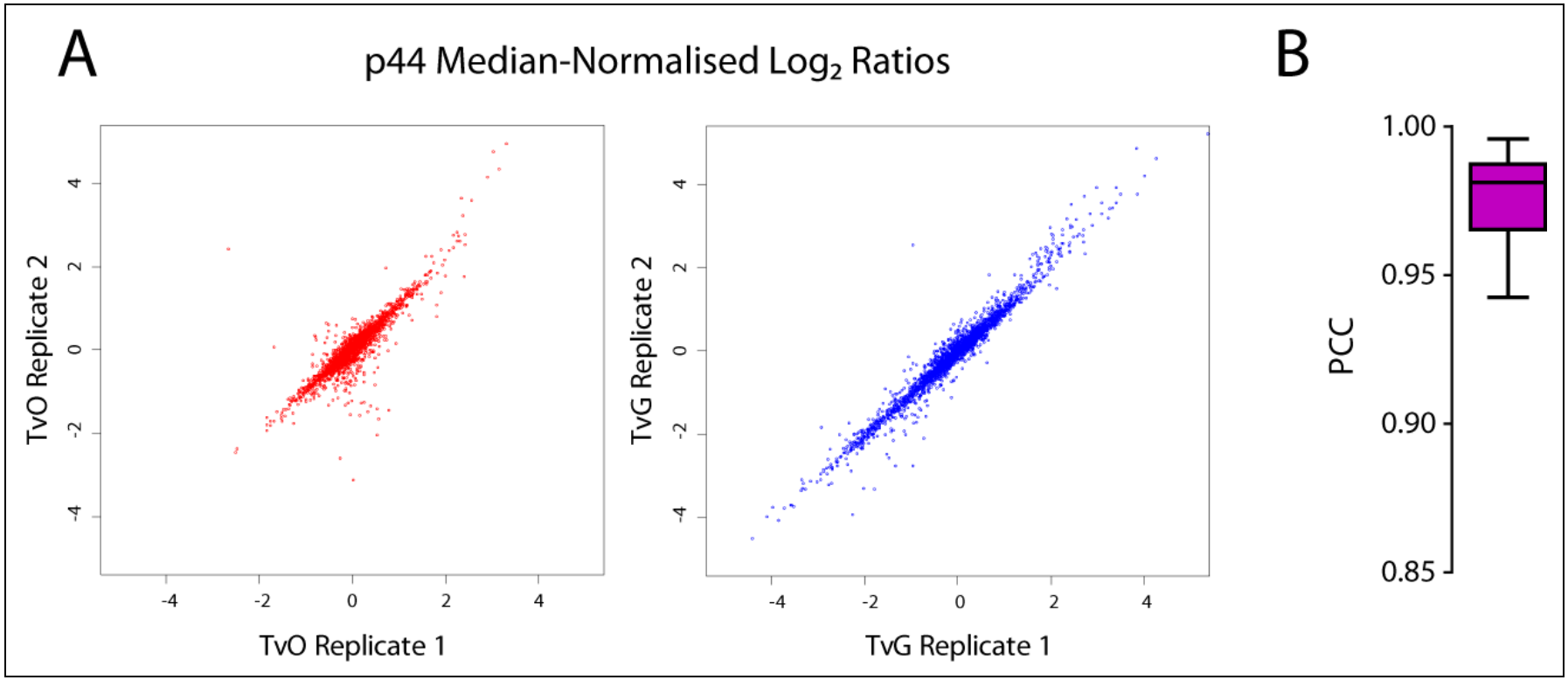

Figure 7-4 Correlation of Technical Replicates. A- Median-normalised log_2 ratios were compared for the two technical replicates of the samples from p44. TvO = adenocarcinoma to normal oesophagus log_2 ratio. TvG = adenocarcinoma to normal gastric log_2 ratio. B – A boxplot summarising the median (line) and 5th and 95th percentiles for the Pearson's correlation coefficients (PCC) for technical replicates across all experiments (n=14).

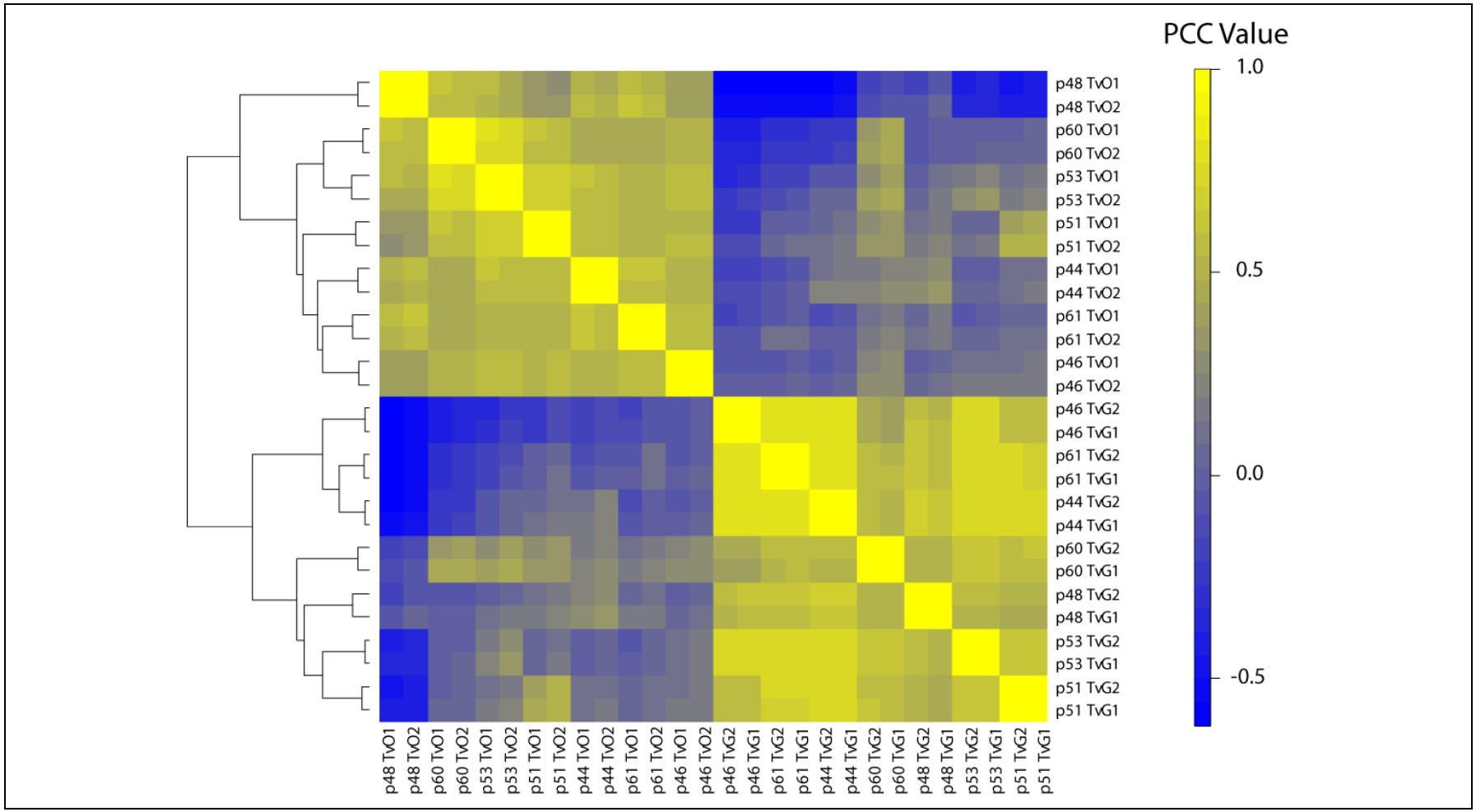

Figure 7-5 Correlation of Normalised Log_2 Tumour : Normal Tissue Ratios. The Pearson's Correlation Coefficient (PCC) was calculated between each experiment and represented in the heat map. Unsupervised hierarchical clustering was performed using an agglomerative complete linkage method to generate the similarity dendrogram (left). This demonstrates the greatest correlation was observed between technical replicate ratios with a clear separation between TvO and TvG ratios.

7.3.3 Tissue-specific protein expression.

To identify tumour specific expression, ratios were compared between multiple tissue types. The analysis method used the ratio of reporter intensities to determine relative expression, yet for proteins where no reporter ion is observed for one or more tissue types, a ratio cannot be calculated. These proteins are potentially tissue specific and therefore all candidates were reviewed where a peptide was detected with reporter expression in only one or two of the patient-matched normal oesophagus, normal gastric and adenocarcinoma samples. In the context of the expression profile of other peptides from the same proteins and other experiments, no proteins could be identified with tissue exclusive expression.

7.3.4 Protein Identifications

Proteins represented by a total of 6,349 different Uniprot accession numbers were quantified in at least one experiment. Due to a small degree of redundancy between TrEMBL and Swissprot accessions in the reference database, these corresponded to proteins from 6,339 unique genes. A total of 744 proteins were quantified in all experiments. The protein identifications per experiment are shown in Table 7-5. Of the 6,339 quantified proteins, 6,221 could be mapped to an Entrez GeneID. The unmapped proteins were either Uniprot entries removed since the reference database release, uncharacterised proteins or immunoglobulins.

Table 7-5 Protein Identifications by Experiment

Patient	Total Proteins (1% peptide FDR, 1 unique peptide)	Unique to sample
P44	2,901	368
P46	2,534	256
P48	3,309	550
P51	3,327	503
P53	2,904	280
P60	2,369	220
P61	2,828	264

7.3.5 Evaluation of Expression Across Experiments

The confidence in the expression value (TvO or TvG ratio) for each protein varies with the number and distribution of peptide ratios contributing to that expression value. A protein may have been quantified in up to 14 experiments across all seven patients' tissues in this study. A mean expression value across studies was therefore sought.

For the 4,181 proteins identified by more than one peptide in more than one experiment, the weighted mean and variance of the $\log_2$ TvO and TvG ratios were calculated across all experiments (see Methods 2.11.3.1). For the proteins only identified by a single peptide, the un-weighted mean and variance were calculated across experiments. Those proteins identified by a single peptide and only in a single experiment were considered low confidence identifications and were excluded from the combined analysis (n=14).

By combining the expression ratios between OAC, normal squamous epithelium and normal gastric epithelium it was hoped proteins expressed generically in glandular tissues would be identified. The two ratios for each protein were used to produce a two-dimensional map with vectors of TvG expression on the x-axis and TvO on the y-axis (Figure 7-6).

7.3.6 Differential Expression

To define differentially expressed proteins, the quantitation thresholds derived from technical replicates could be applied (Table 7-4) and proteins outside those limits accepted as dysregulated. When considering several thousand protein expression values, however, there is considerable potential for a type I statistical error due to multiple hypothesis testing. As the accuracy of quantitation also varied from protein to protein, a protein-wise calculation of significance with a false discovery rate correction was sought.

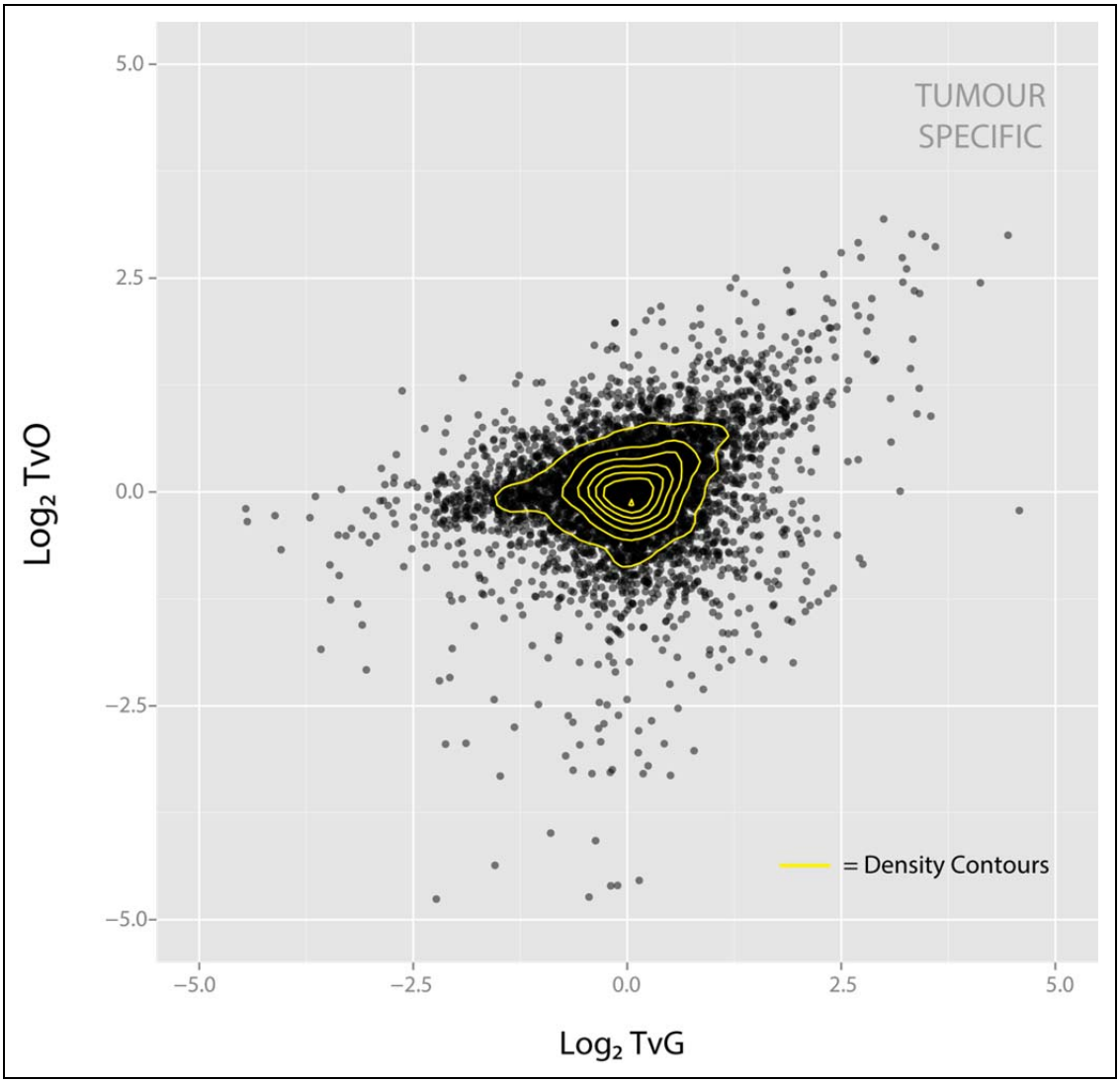

Figure 7-6 Protein Expression According to Tissue Type. Each point represents a protein defined by Log2 Tumour vs. Normal Oesophageal (TvO) and Tumour vs. Gastric (TvG) ratios. The density contours describe the distribution of over-plotted points surrounding the origin

For each protein, the ratios for each tissue comparison were derived from a variable number of experiments and, for each experiment, from a variable number of peptides. The variance in the mean ratio and the number of peptides contributing to the mean were used to test the hypothesis that, within the experimental data observed, the TvO or TvG ratio for that protein was not significantly different to the ratio produced from the comparison of technical replicates (TvT, OvO and GvG). The probability associated with a statistical test of this hypothesis was denoted as the "intra-experimental probability".

It was predicted that given the relative accuracy of TMT-based quantitation, the large number of peptides assessed, and the histological differences between tissue types, many proteins would show significantly different expression between tissues. Many proteins were quantified in less than half of the experiments, however, and the degree to which the confidently determined expression values simply represent normal fluctuation between biological samples is much less certain (Figure 7-7).

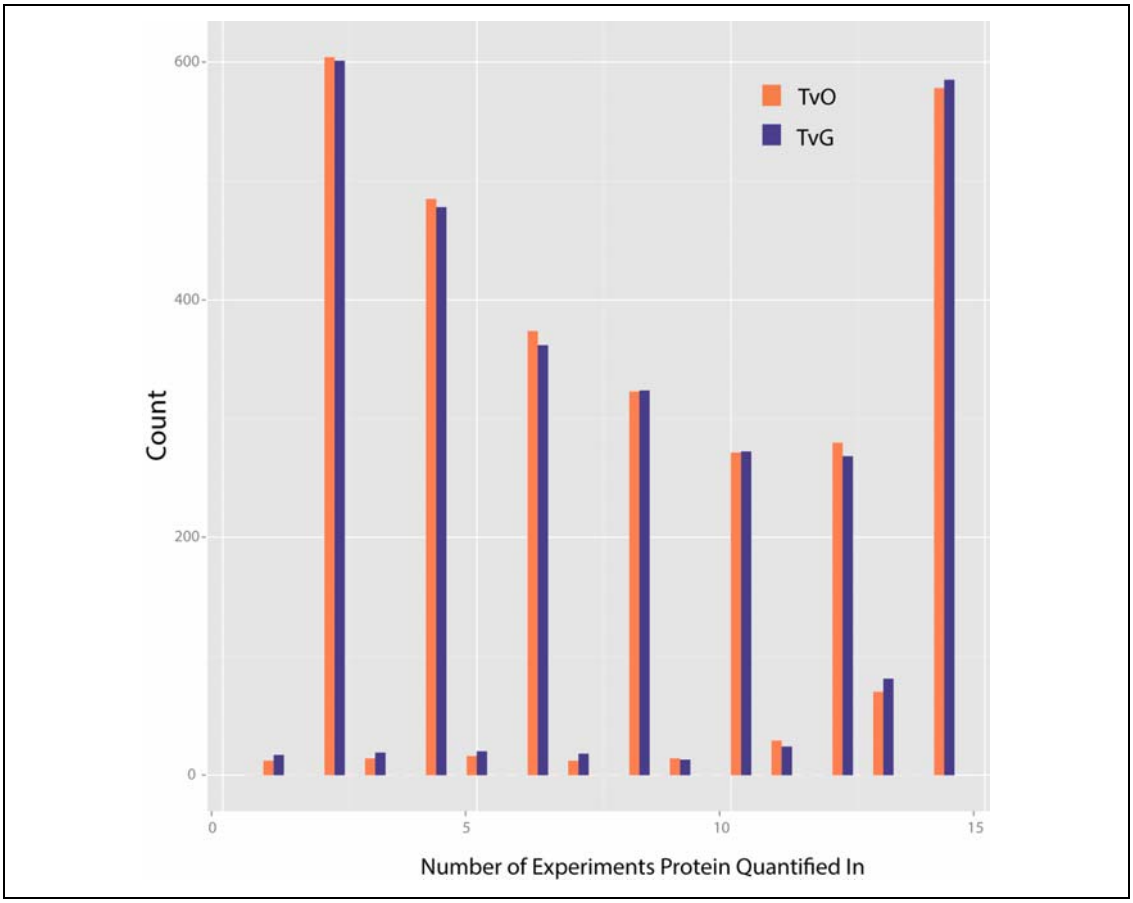

Figure 7-7 Frequency of Differentially Expressed Protein Quantitation Across Experiments. Only proteins with an intra-experimental FDR-corrected p<0.05 were considered.

To predict which proteins may be dysregulated in all oesophageal adenocarcinomas when compared to normal squamous and gastric epithelium it was assumed the 7 patients tissues were representative, random samples of the total population and the population protein abundance was normally distributed. The probability that a protein was not differentially expressed was calculated from the comparison of TvO or TvG ratios and technical replicate ratios using the number of patients as the "n". This probability was termed the "inter-experimental probability".

A potential method to calculate the intra and inter-experimental p-value for each protein would be to apply the Analysis of Variance (ANOVA) test across TvO, TvG and technical replicate ratios. This would determine if proteins showed significantly different expression between the ratios and post-hoc tests could identify which comparisons were significant. This method assumes covariates are normally distributed and have an equal variance. The distributions of protein ratios is summarised in Figure 7-8A.

Formal testing of deviation from a normal distribution can be accomplished with Shapiro-Wilks or Lillefors tests. Both these tests revealed a significant deviation from a normal distribution for all the ratios ($p<10^{-10}$ for all ratios, both tests). Many described tests of deviation from normality exhibit high sensitivity when a high sample number is used, however, and even small deviations from the theoretical normal distribution can be identified. The sensitivity of these tests was verified when randomly selected, normally distributed variables containing 6,349 numbers with the same mean and standard deviation as the experimentally determined ratios were subjected to normality testing (Table 7-6). As these variable were selected to fit a normal distribution using computational methods, the predicted percentage of p-values <0.05 from the normality tests should be <5%. Tests prone to false positives would demonstrate a higher proportion than this. Both tests demonstrated over-sensitivity across all ratios (Table 7-6).

Table 7-6 Percentage of incorrect results (p<0.05) from normality tests using simulated normally-distributed data.

	Ratio Determining the Mean and SD of the Simulated Variable				
	TvO	TvG	TvT	OvO	GvG
Shapiro-Wilks	64.9%	61.4%	38.6%	31.6%	37.4%
Lillefors	51.3%	26.3%	20.1%	41.7%	18.4%

From closer examination of the histograms of expression ratios, it appears the distributions are only approximately normal with some evidence of leptokurtosis (Figure 7-2, Figure 7-8A). The variance of leptokurtotic distributions is overestimated using standard calculations and this could lead to a conservative estimate of the number of dysregulated proteins. ANOVA has, however, been shown to be relatively robust to small deviations from normality if the sample variances are approximately equal[684].

A global test for the equivalence of sample variances, the Fligner-Killeen test, demonstrated significant differences between sample variances ($p<10^{-10}$). This can be seen graphically in Figure 7-8B where there is a clear difference in the variance between the technical replicates and the TvO and TvG ratios.

The combination of deviation from normality and heteroscedascity make ANOVA testing inappropriate. Welch's test is a parametric statistic that can be applied to compare two means from approximately normally distributed populations with unequal variances to test the hypothesis that the means are the same[512]. Similar to ANOVA, this statistic is also relatively robust to deviation from normality when the sample size is large. Welch's test was therefore applied to compare the mean TvO ratio to the global mean of TvT and OvO ratios or similarly the mean TvG ratio to the global mean of TvT and GvG ratios for each quantified protein. In this fashion both intra and inter-experimental p-values were calculated for all proteins with the false-discovery rate controlled using the Benjamini Yekutieli method[566].

7.3.7 Evaluation of Inter-Experimentally Dysregulated Proteins

Proteins commonly dysregulated across experiments were considered more compelling targets for therapy as they may be relevant to a larger proportion of patients. The proteins with an inter-experimental FDR-corrected $p < 0.05$ for TvO ratios are demonstrated in Figure 7-9. Similarly proteins with an inter-experimental FDR-corrected $p<0.05$ for TvG ratios are demonstrated in Figure 7-10. No proteins showed an inter-experimental FDR-corrected p-value <0.05 for both TvO and TvG ratios. This highlights the limitations imposed by the small sample size on the power to predict dysregulated proteins in OAC using the inter-experimental p-value alone.

A more liberal, and potentially more informative approach, is to use the intra-experimental FDR-corrected p-values combined with ratio thresholds to define significantly dysregulated proteins. The outliers identified using this approach may simply represent random biological variation, however, by also incorporating prior knowledge such as previous studies of protein expression in OAC or known roles in cancer, compelling candidates can be shortlisted for validation.

7.3.8 Evaluation of Intra-Experimentally Dysregulated Proteins

For intra-experimental comparisons, proteins with an FDR-corrected p-value >0.05 for either TvG or TvO ratios were removed leaving 3,082 significantly dysregulated proteins. The distribution of these proteins is plotted in Figure 7-11 .

When the candidates previously examined by IHC are superimposed, it can be seen that the TvG dimension assists in the discrimination of proteins expressed in both OAC and gastric epithelium (Figure 7-12). From this 2D map, it becomes clear that certain areas, defined by a specific expression profile are populated with proteins that appear to share similar properties such as tissue expression. To further assess the utility of this expression map to reveal tissue specific expression, several proteins with well validated tissue expression profiles were selected. Their distribution on the expression map highlights regions associated with gastric and normal squamous oesophageal specific expression (Figure 7-12).

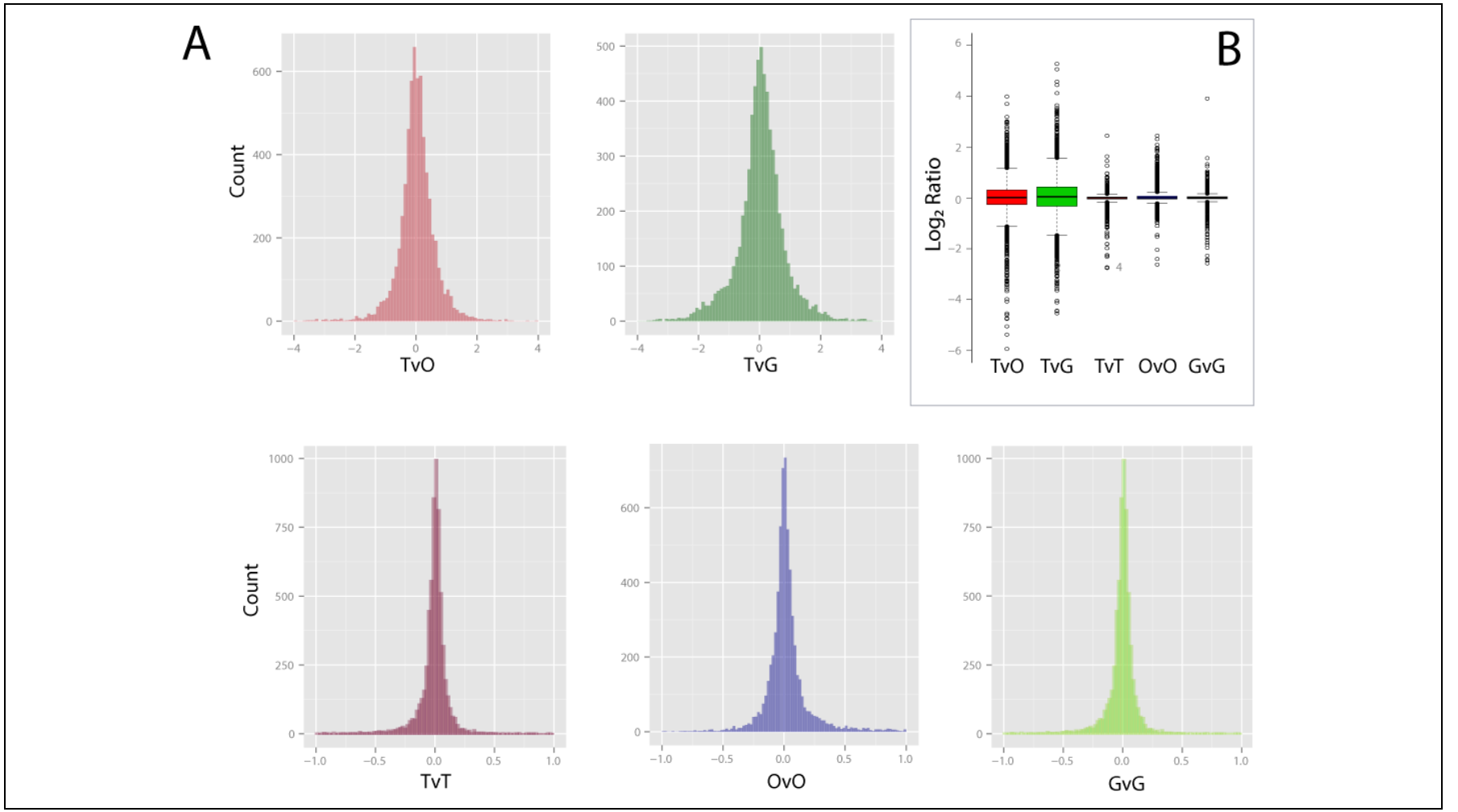

Figure 7-8 (A) Histograms showing the distribution of normalised Log_2 ratios. Note the differing x-axis (log) scales. (B) Boxplots of the same data. Boxplot whiskers represent 1.5 Interquartile ranges above or below the 75th and 25th percentiles respectively. Unfilled circles represent outlier points.

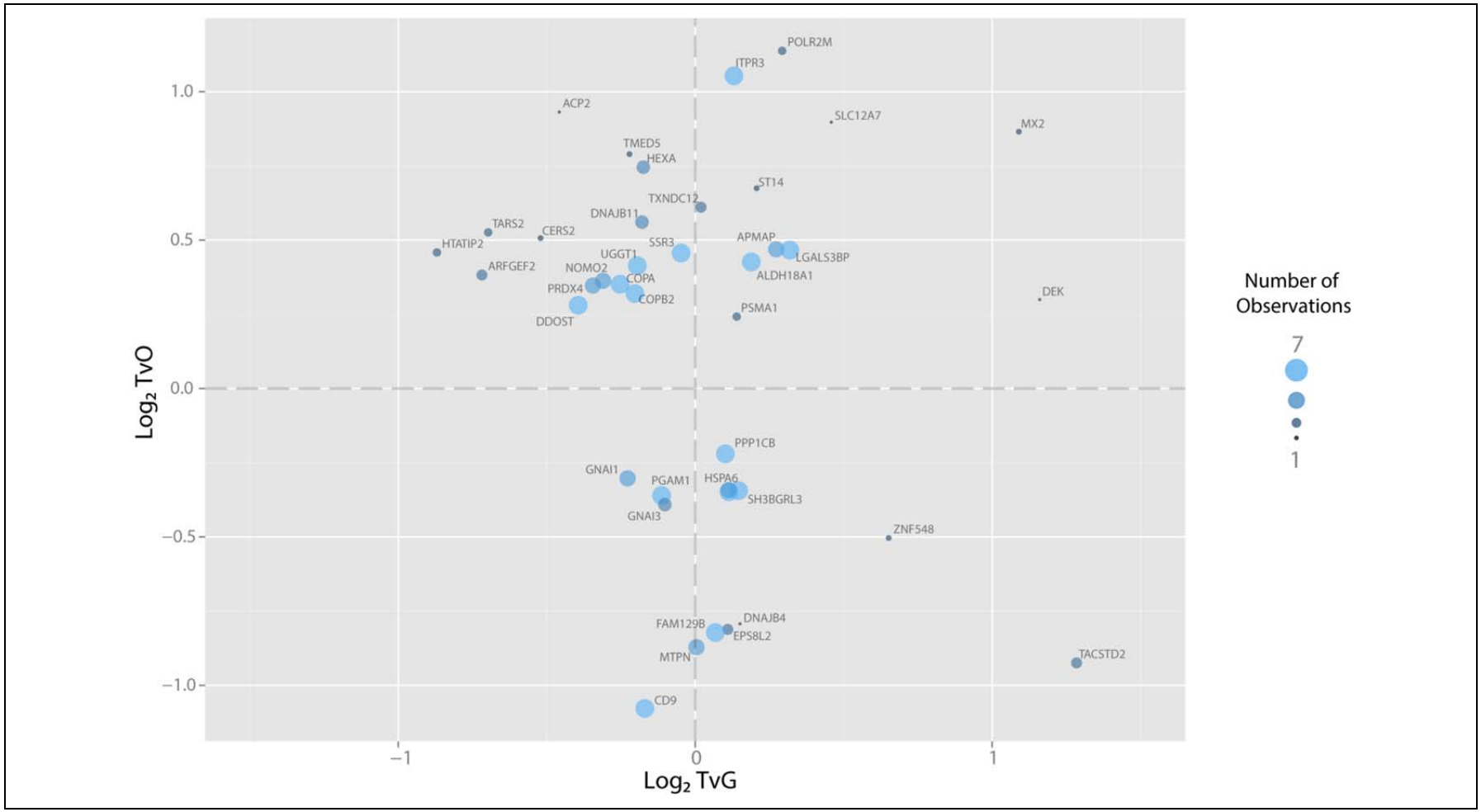

Figure 7-9 Proteins Significantly Dysregulated between Tumour and Normal Oesophageal Tissue. Significance was defined as an inter-experimental FDR-corrected $p<0.05$ for the comparison of TvO and the mean of TvT and OvO ratios. Proteins have been labelled with their gene names and plotted according to their log_2 expression ratios.

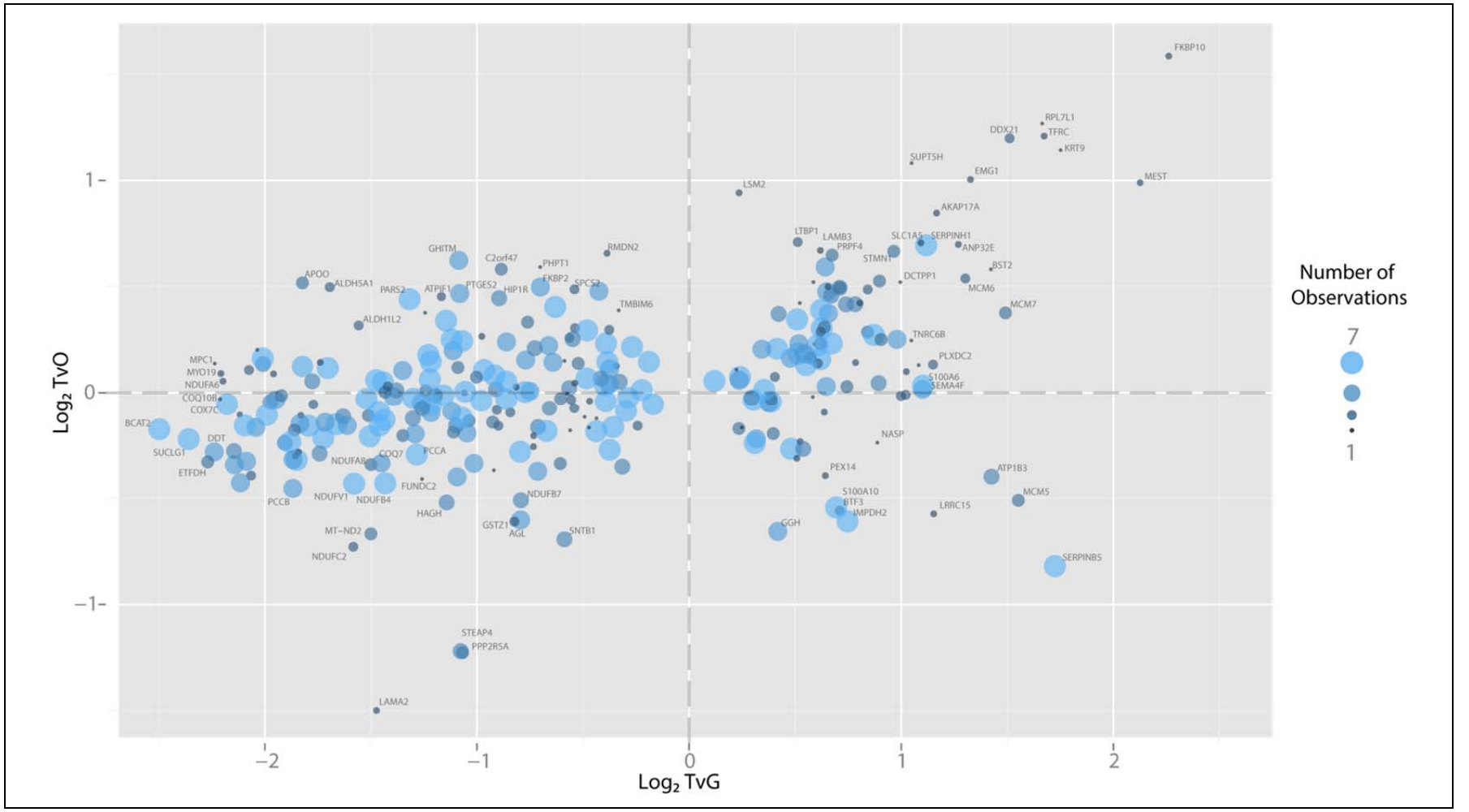

Figure 7-10 Proteins Significantly Dysregulated between Tumour and Normal Gastric Tissue. Significance was defined as an inter-experimental FDR-corrected $p<0.05$ for the comparison of TvG and the mean of TvT and GvG ratios. Proteins have been labelled with their gene names and plotted according to their log_2 expression ratios.

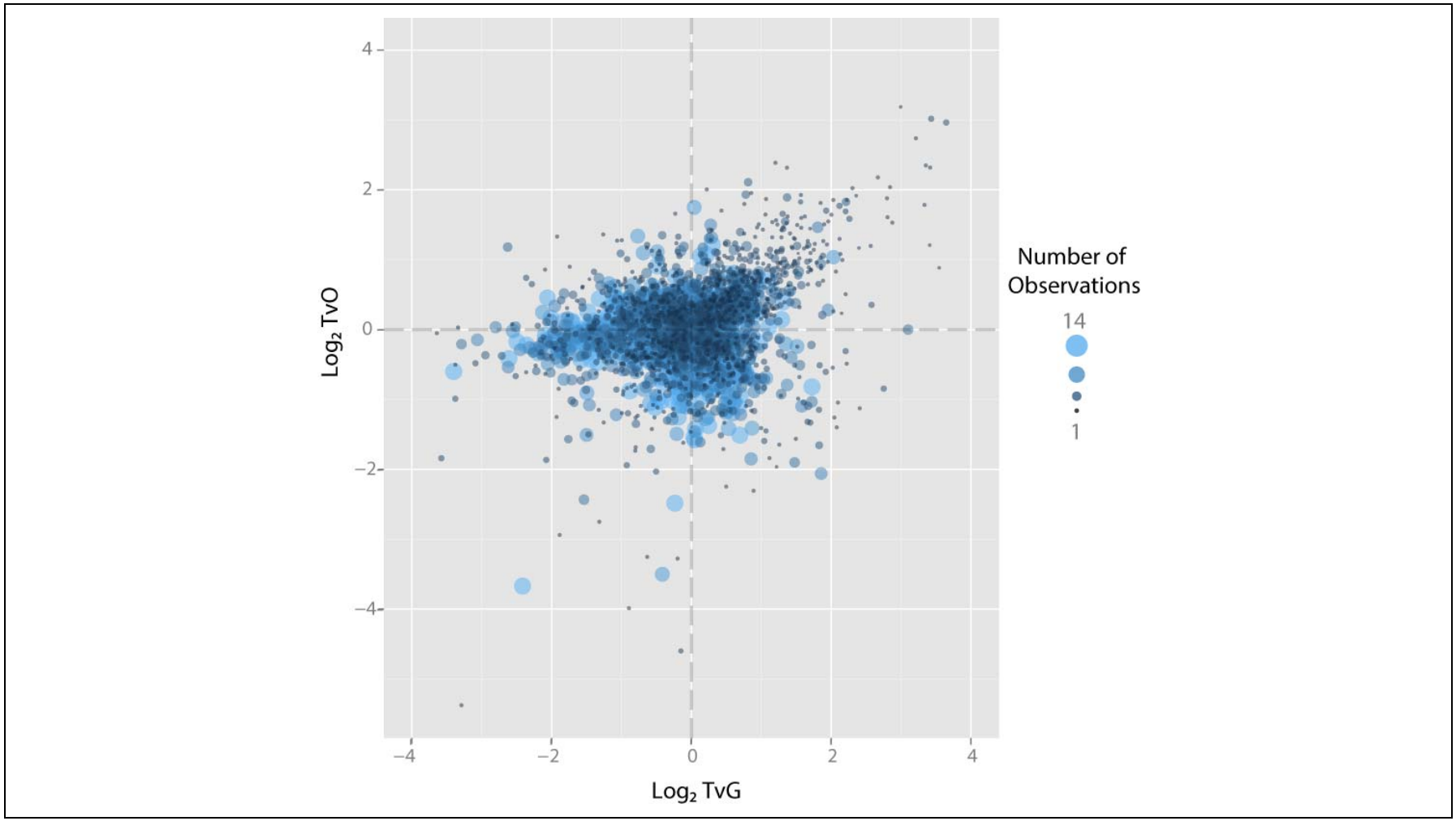

Figure 7-11 Distribution of Significantly Dysregulated Proteins. Significance was defined as intra-experimental FDR-corrected p<0.05 for the comparison of TvO and the mean of TvT and OvO ratios and for the comparison of TvG and the mean of TvT and GvG ratios. Proteins have been plotted according to their log_2 expression ratios and points sized according to the number of experimental observations.

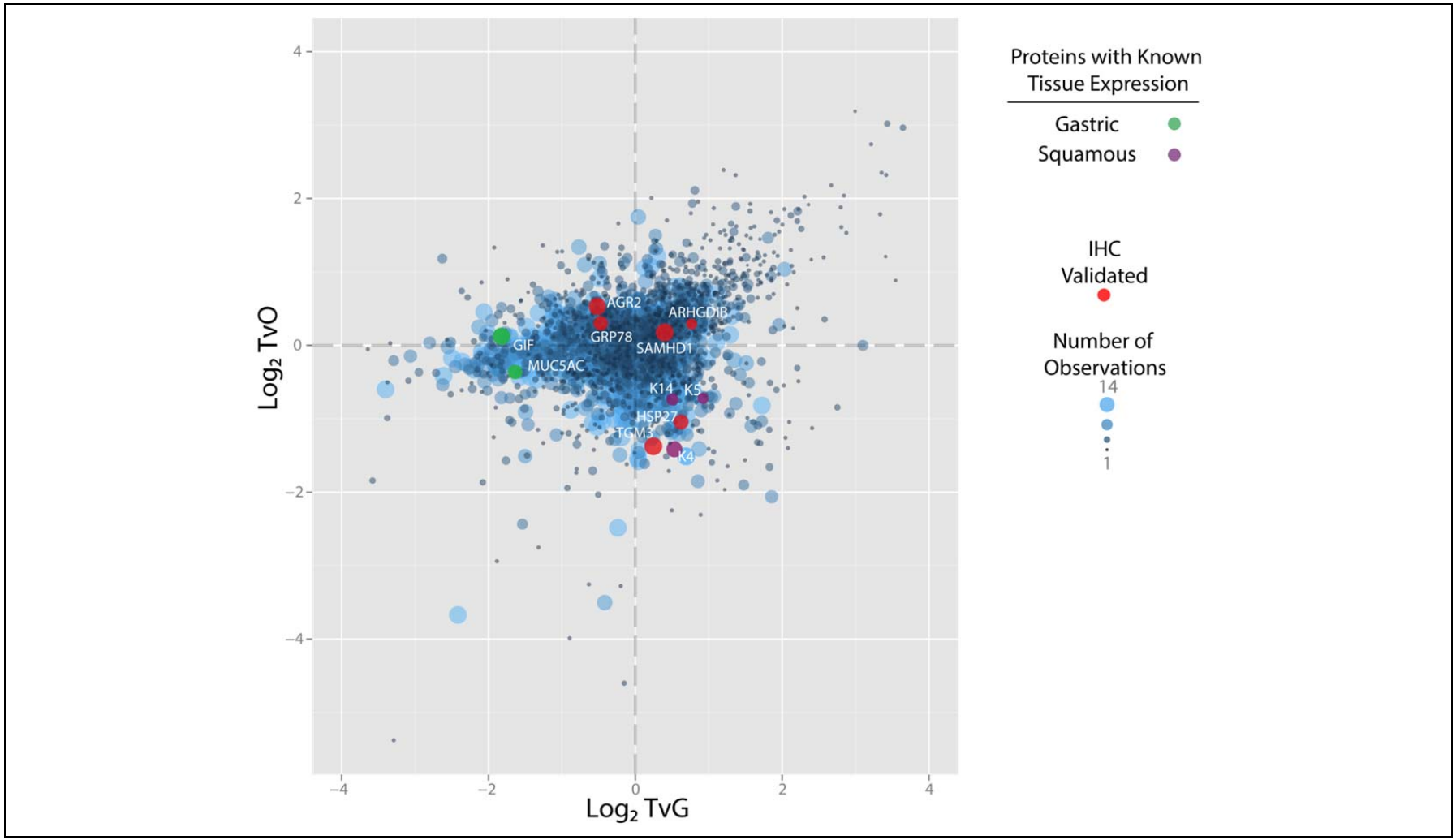

Figure 7-12 The Expression of Proteins Previously Validated by IHC and Selected Examples with Known Tissue Expression. Points are identical to Figure 7-11 but have been highlighted in selected cases according to published tissue specific expression or previous validation of tissue expression by IHC and annotated with gene names.

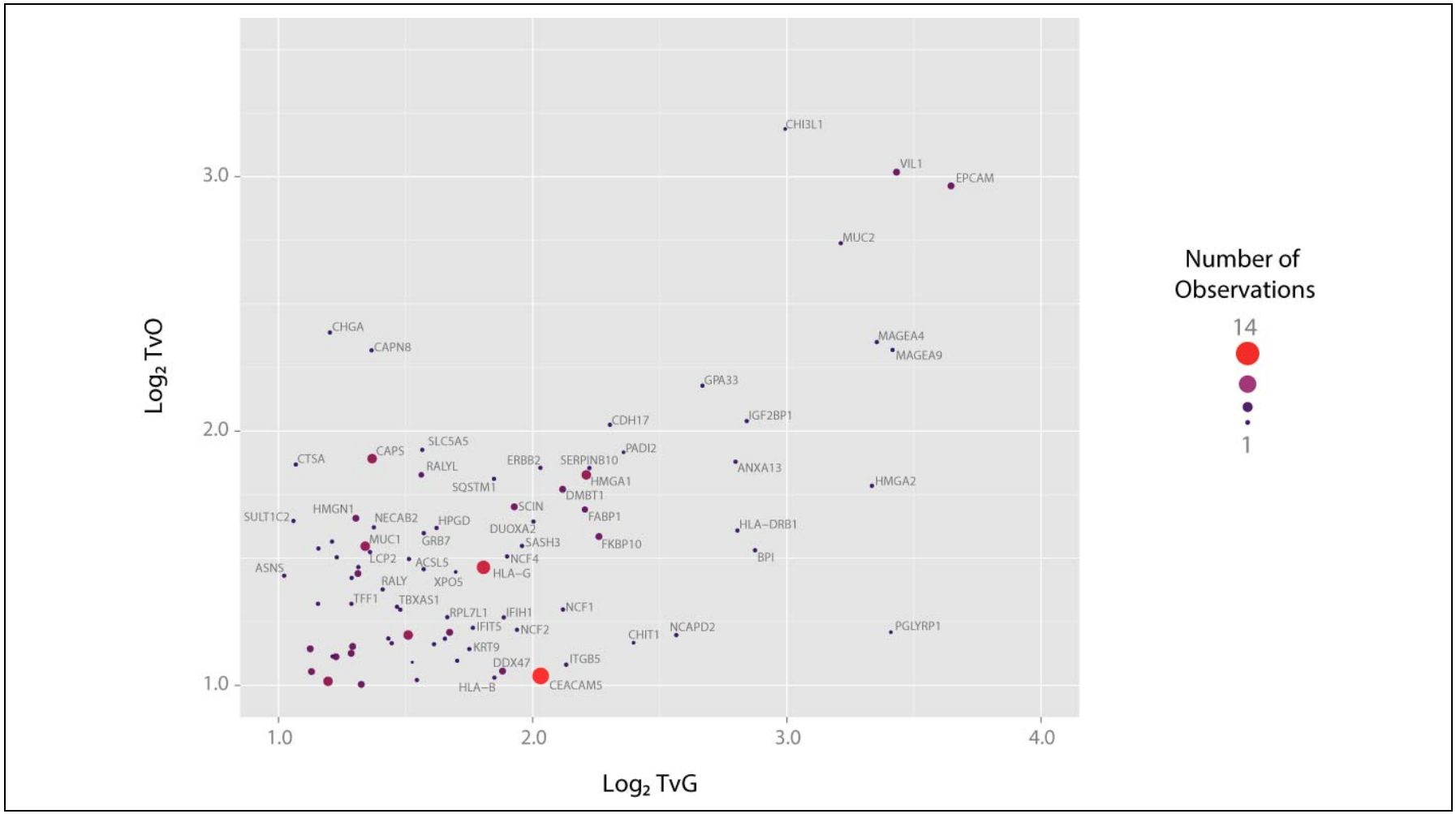

Figure 7-13 Annotation of Proteins Highly Expressed in Tumours Compared to Normal Oesophageal and Gastric Epithelium. Points are identical to Figure 7-11 but only proteins within a selected expression range are shown. Proteins have been annotated with gene names and are sized according to the number of experimental observations.

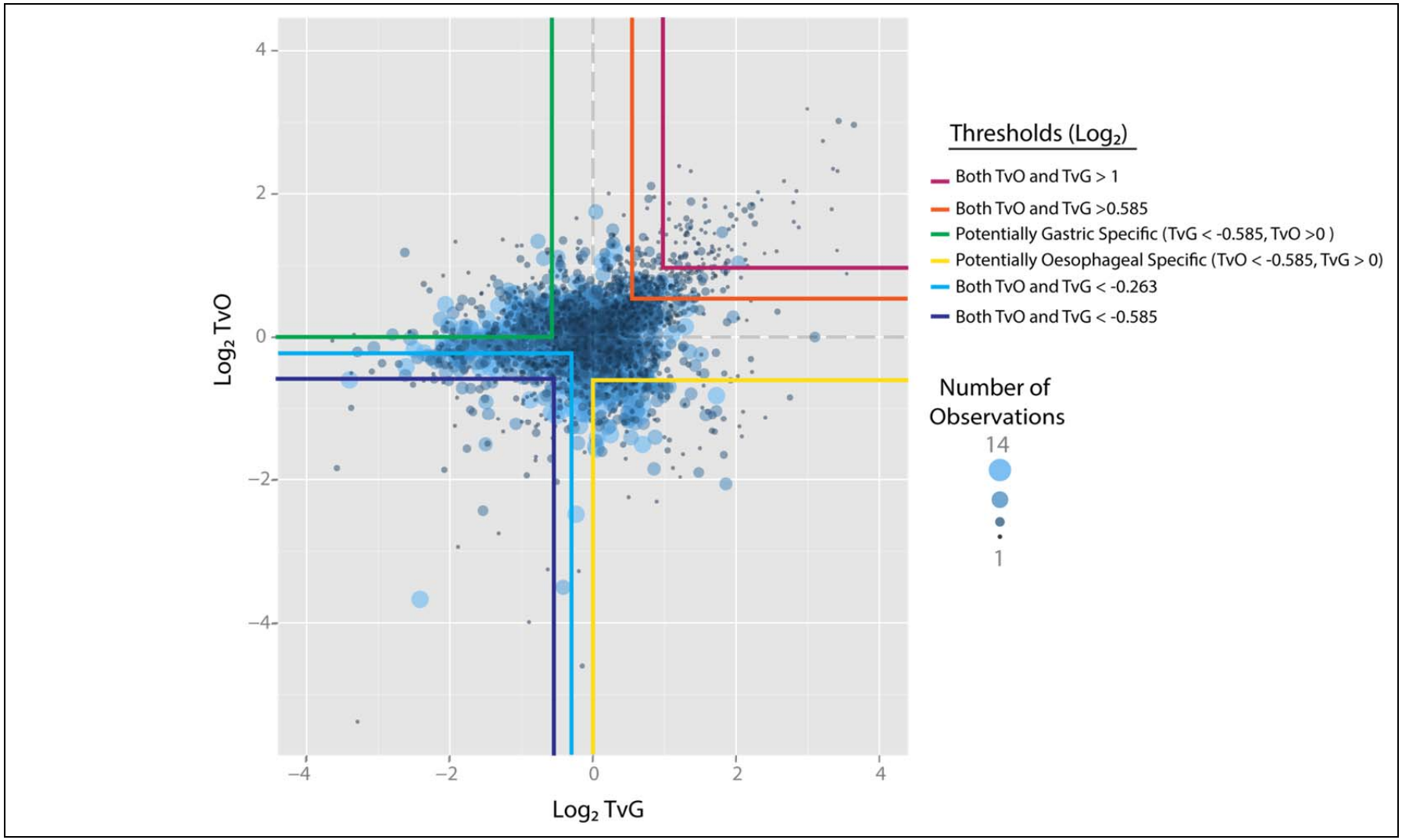

Figure 7-14 Summary of expression thresholds used to define groups of proteins for network analysis. Thresholds have been superimposed onto Figure 7-11.

Closer annotation of the proteins with high expression in tumours relative to both normal oesophageal and gastric epithelium reveals several compelling potential therapeutic targets (Figure 7-13, see discussion). A number of these were identified in the candidate tumour specific list derived from the PAcIFIC data using the Bayesian method (Table 5-14).

It was predicted that proteins with particular tissue expression profiles would be located in discrete regions on the 2D map. Thresholds were created to select proteins with specific expression profiles (Figure 7-14). To further evaluate the function of proteins within these provisional tissue specific expression areas, a network analysis was undertaken. Groups of proteins, defined by the expression thresholds, were mapped to the Linghu network[460] as previously described (Methods 2.13).

7.3.9 Network Analysis

Six lists of significantly dysregulated genes, with potentially tissue-specific functions were mapped to the LR5+ Linghu FLN (Figure 7-14). Only genes with an intra-experimental FDR-corrected $p<0.05$ were included. The number of significant clusters ($p<0.1$) identified for each list is demonstrated Table 7-7. An inflation value (I-value) of 5 was selected as providing the optimal balance between the number of clusters and the network complexity. The biological process ontologies for the significantly enriched clusters are shown in Table 7-8. Network visualisations are provided in Figure 7-15 to Figure 7-20.

Table 7-7 Number of Significant Clusters Identified Across MCL Inflation-Values using the mFASP TMT data

Number of Significantly Enriched Clusters (FDR-corrected p<0.1)									
Differentially Expressed List	Inflation Value								
Combined mFASP TMT Intra-FDR p<0.05	2	2.5	3	3.5	4	4.5	5	5.5	6
Both TvO and TvG 1.5 fold up-regulated	5	5	5	4	6	5	5	4	5
Both TvO and TvG 2 fold up-regulated	3	1	2	2	3	3	3	2	3
Both TvO and TvG 1.2 fold down-regulated	7	0	6	4	7	6	2	7	0
Both TvO and TvG 1.5 fold down-regulated	0	1	1	1	4	1	3	0	0
Gastric Specific	0	4	3	1	1	2	2	2	2
Oesophageal Specific	3	4	5	5	6	6	6	0	0

Table 7-8 Significant Clusters from the mFASP TMT Dysregulated Protein Lists Mapped onto the Linghu FLN clustered at an MCL Inflation value of 5.0.

mFASP TMT Oesophageal-Specific List							
Cluster No	Mapped Genes	Quantified Genes	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment FDR-corrected p-values	Summary Gene Class	Cluster FDR-corrected p-value
9	5	31	108	[Calcium ion Transport] [Cation transport] [Metal Ion transport]	[$1.7x10^{-5}$] [$2.6x10^{-35}$] [$3x10^{-3}$]	S100 and Calcium Binding proteins	0.02733
662	3	4	4	[Cell adhesion]	[$1.4x10^{-3}$]	Plakophilins	0.02733
41	4	19	34	[Regulation of Proteolysis]	[$3x10^{-6}$]	Serpin Peptidase Inhibitors	0.02733
13	14	29	70	[Ectoderm Development] [Epidermis Development] [Intermediate Filament-based Process]	[$5.3x10^{-22}$] [$3.2x10^{-19}$] [$5.7x10^{-17}$]	Keratins	0.00205
42	5	18	31	[Muscle Contraction] [Cytoskeletal Organisation] [Actin filament-based process]	[$6.9x10^{-13}$] [$8.1x10^{-5}$] [$2.5x10^{-4}$]	Cytoskeletal Proteins	0.002733
184	5	6	11	[Keratinisation] [Keratinocyte Differentiation] [Epidermal Cell Differentiation]	$7.5x10^{-25}$] [$4.1x10^{-23}$] [$7.0x10^{-23}$]	Small proline rich proteins	0.00205
mFASP TMT Gastric-Specific List							
Cluster No	Mapped Genes	Quantified Genes	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment FDR-corrected p-values	Summary Gene Class	Cluster FDR-corrected p-value
53	6	23	29	[Electron transport chain] [Mitochondrial electron transport, NADH to ubiquinone] [ATP synthesis coupled electron transport]	[$2.6x10^{-45}$] [$1.3x10^{-45}$] [$8x10^{-43}$]	NADH Dehydrogenases	0.0092
167	4	11	13	[Mitochondrial electron transport, Ubiquinol to cytochrome c] [cellular respiration] [Electron transport chain]	[$5.6x10^{-11}$] [$3x10^{-10}$] [$5.3x10^{-10}$]	Aldehyde Dehydrogenases	0.0575
mFASP TMT both TvG and TvO 1.2 fold Down-regulated							
Cluster No	Mapped Genes	Quantified Genes	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment FDR-corrected p-values	Summary Gene Class	Cluster FDR-corrected p-value
53	12	23	29	[Electron transport chain] [Mitochondrial electron transport, NADH to ubiquinone] [ATP synthesis coupled electron transport]	[$2.6x10^{-45}$] [$1.3x10^{-45}$] [$8x10^{-43}$]	NADH Dehydrogenases	0.0026
232	6	9	10	[Oxidation Reduction] [Cellular Aldehyde Metabolic Process]	[$9.9x10^{-11}$] [$1.5x10^{-10}$]	Aldehyde Dehydrogenases	0.0026

Table 7-9 Significant Clusters from the TMT Dysregulated Lists Mapped onto the Linghu Network clustered at an MCL Inflation value of 5.0 (continued).

mFASP TMT both TvG and TvO 1.5 fold Down-regulated							
Cluster No	Mapped Genes	Quantified Genes	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment FDR-corrected p-values	Summary Gene Class	Cluster FDR-corrected p-value
16	3	31	60	[Cell Adhesion] [Biological Adhesion] [Regulation of cell migration]	[$1.6x10^{-23}$] [$8.1x10^{-24}$] [$7.3x10^{-5}$]	Laminin and ECM proteins	0.06333
53	3	23	29	[Electron transport chain] [Mitochondrial electron transport, NADH to ubiquinone] [ATP synthesis]	[$2.6x10^{-45}$] [$1.3x10^{-45}$] [$8x10^{-43}$]	NADH Dehydrogenases	0.03625
232	3	9	10	[Oxidation Reduction] [Cellular Aldehyde Metabolic Process]	[$9.9x10^{-11}$] [$1.5x10^{-10}$]	Aldehyde Dehydrogenases	0.019
mFASP TMT both TvG and TvO 1.5 fold Up-regulated							
Cluster No	Mapped Genes	Quantified Genes	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment FDR-corrected p-values	Summary Gene Class	Cluster FDR-corrected p-value
7	10	63	140	[RNA processing] [mRNA processing] [RNA splicing]	[$4.0x10^{-77}$] [$1.2x10^{-75}$] [$1.7x10^{-55}$]	RNA binding proteins	0.033
16	6	31	60	[Cell Adhesion] [Biological Adhesion] [Regulation of cell migration]	[$1.6x10^{-23}$] [$8.1x10^{-24}$] [$7.3x10^{-5}$]	Laminin and ECM proteins	0.03437
26	4	13	39	[Antigen Presentation] [Immune Response] [Antigen processing via MHC class II]	[$2.9x10^{-44}$] [$6.0x10^{-32}$]	HLA and T-cell receptor genes	0.034375
70	4	5	24	[Immune Response] [Antigen Presentation] [Antigen processing via MHC class I]	[$2.3x10^{-18}$] [$1.3x10^{-15}$] [$1.6x10^{-13}$]	HLA receptor genes	0.0451
82	3	12	20	[RNA metabolic process]	[$3.9x10^{-2}$]	DEAD box proteins	0.03267
mFASP TMT both TvG and TvO 2 fold Upregulated							
Cluster No	Mapped Genes	Quantified Genes	Cluster Size (Genes)	Gene Ontologies Top 3 Biological Processes	GO enrichment FDR-corrected p-values	Summary Gene Class	Cluster FDR-corrected p-value
26	4	13	39	[Antigen Presentation] [Immune Response] [Antigen processing (MHC class II)]	[$2.9x10^{-44}$] [$6.0x10^{-32}$]	HLA and T-cell receptor genes	0.0035
70	3	5	24	[Immune Response] [Antigen Presentation] [Antigen processing via MHC class I]	[$2.3x10^{-18}$] [$1.3x10^{-15}$] [$1.6x10^{-13}$]	HLA receptor genes	0.03267
82	3	12	20	[RNA metabolic process]	[$3.9x10^{-2}$]	DEAD box proteins	0.03267

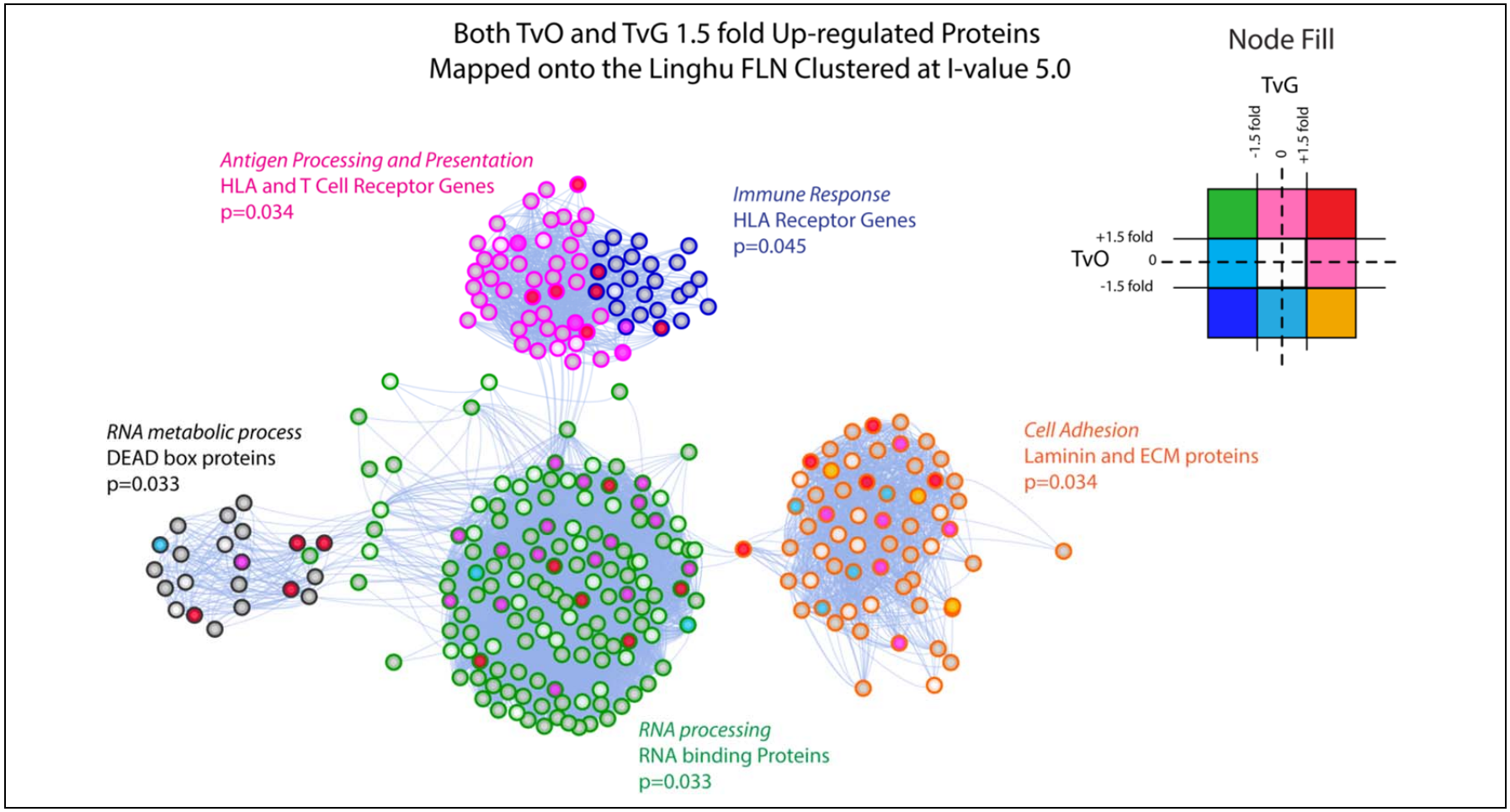

Figure 7-15 Significantly Enriched Functional Clusters in the Genes with both TvG and TvO > 1.5. The top biological process ontology for each cluster is detailed in italics, a summary of the cluster functional class and the FDR-corrected p-value for the enrichment of the cluster are displayed. Nodes have been filled according to TvO and TvG expression so that proteins up-regulated > 1.5 fold in tumours compared to normal oesophagus have been filled red, pink or green depending on their TvG ratio and so on. Node outlines correspond to cluster membership.

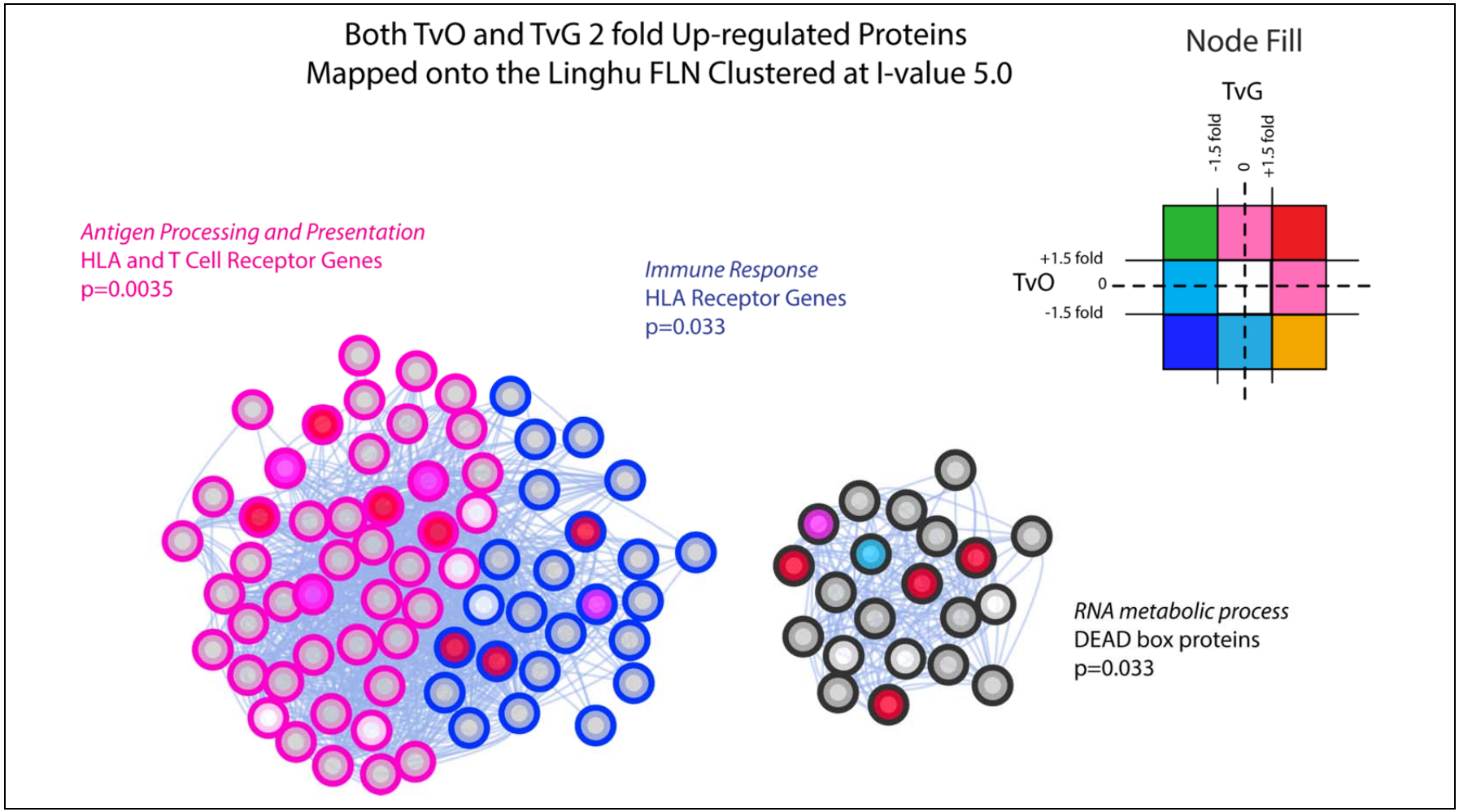

Figure 7-16 Significantly Enriched Functional Clusters in the Genes with both TvG and TvO > 2. The top biological process ontology for each cluster is detailed in italics, a summary of the cluster functional class and the FDR-corrected p-value for the enrichment of the cluster are displayed. Nodes have been filled according to TvO and TvG expression as in Figure 7-15 with node outlines corresponding to cluster membership.

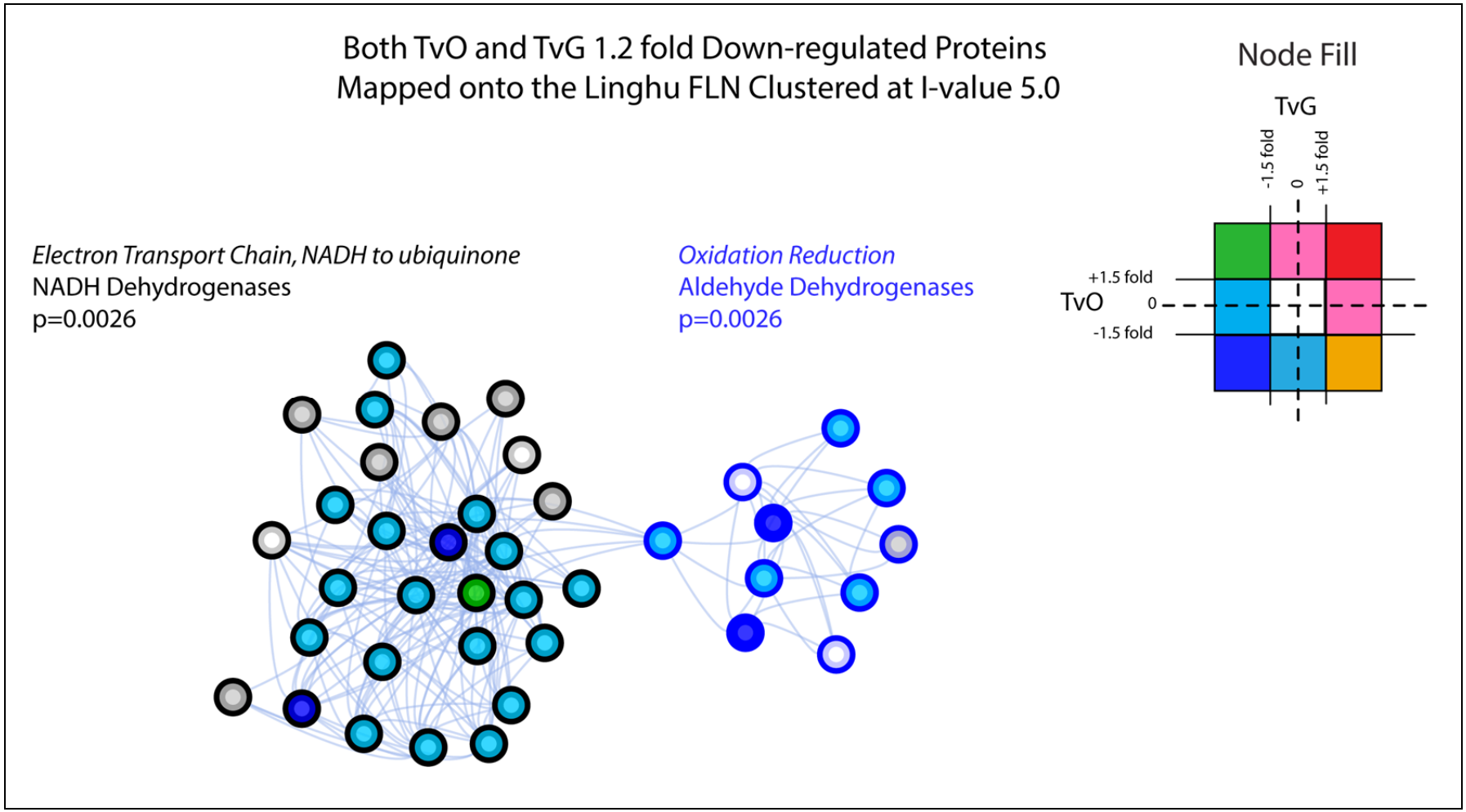

Figure 7-17 Significantly Enriched Functional Clusters in the Genes1.2 fold down-regulated in both TvO and TvG. The top biological process ontology for each cluster is detailed in italics, and a summary of the cluster functional class and the FDR-corrected p-value for the enrichment of the cluster are displayed. Nodes have been filled according to TvO and TvG expression as in Figure 7-15 with node outlines corresponding to cluster membership.

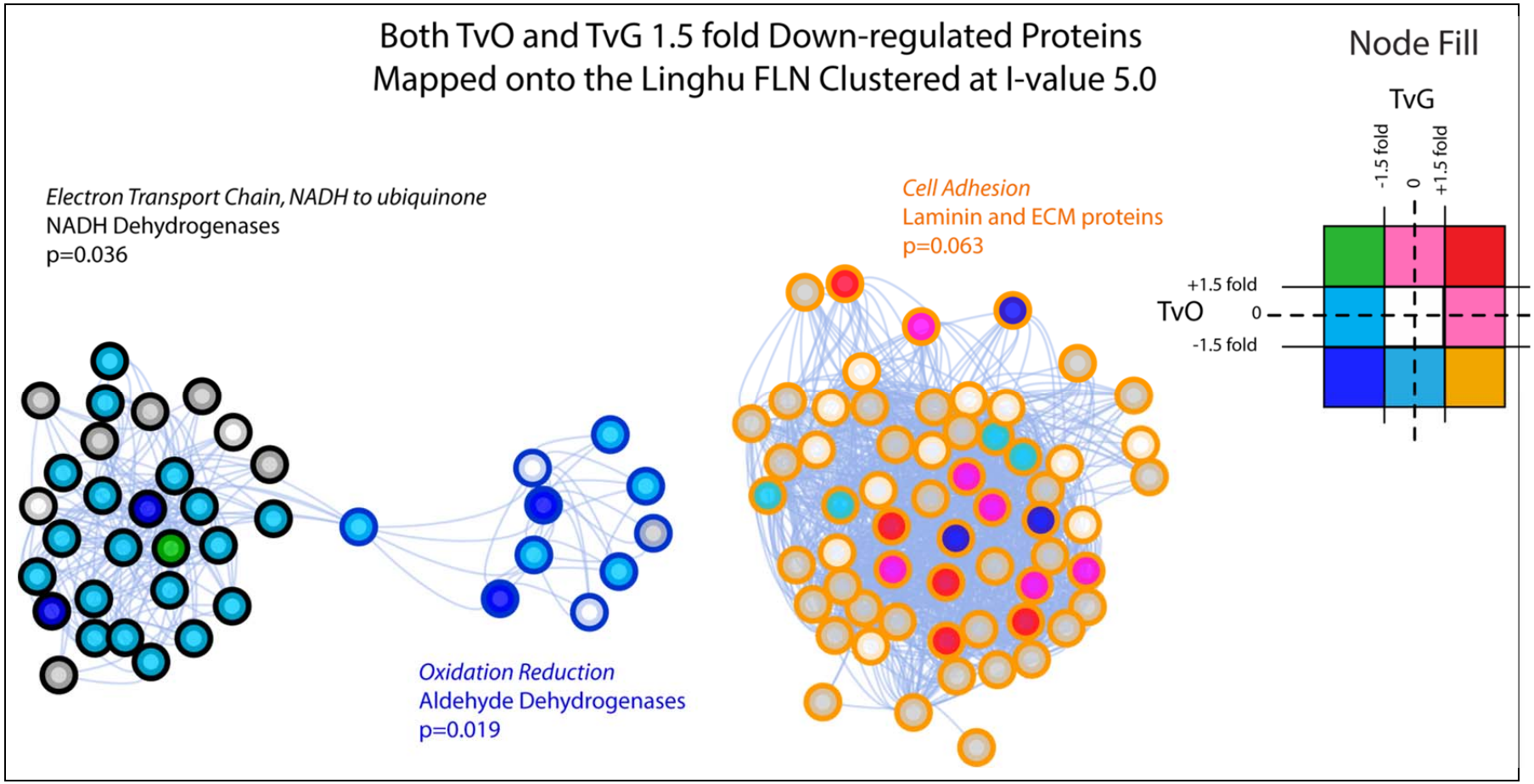

Figure 7-18 Significantly Enriched Functional Clusters in the Genes 1.5 fold Down-regulated in both TvO and TvG. The top biological process ontology for each cluster is detailed in italics, a summary of the cluster functional class and the FDR-corrected p-value for the enrichment of the cluster are displayed. Nodes have been filled according to TvO and TvG expression as in Figure 7-15 with node outlines corresponding to cluster membership.

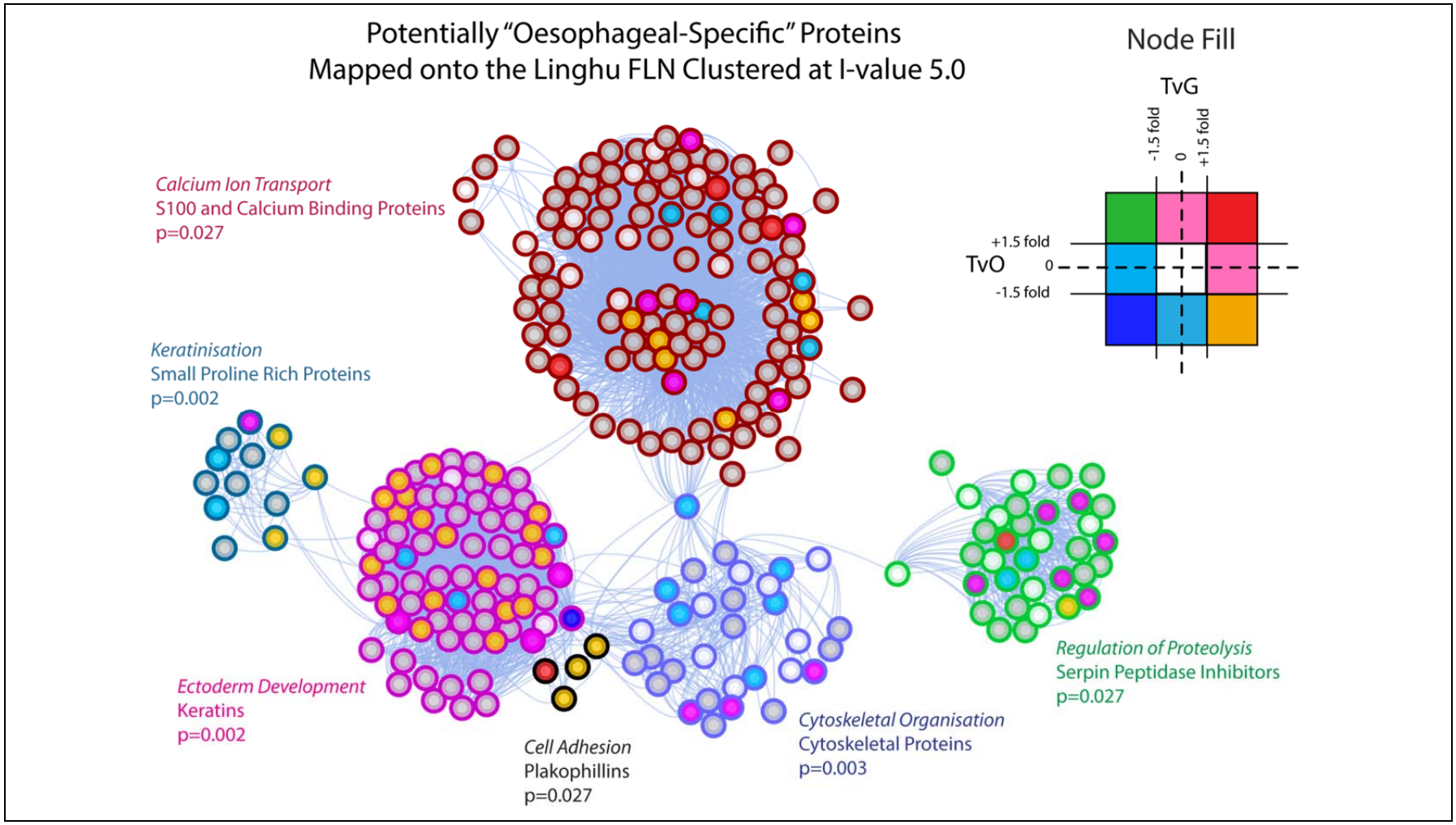

Figure 7-19 Significantly Enriched Functional Clusters in the "Oesophageal Specific" Genes. The top biological process ontology for each cluster is detailed in italics, and a summary of the cluster functional class and the FDR-corrected p-value for the enrichment of the cluster are displayed. Nodes have been filled according to TvO and TvG expression as in Figure 7-15 with node outlines corresponding to cluster membership.

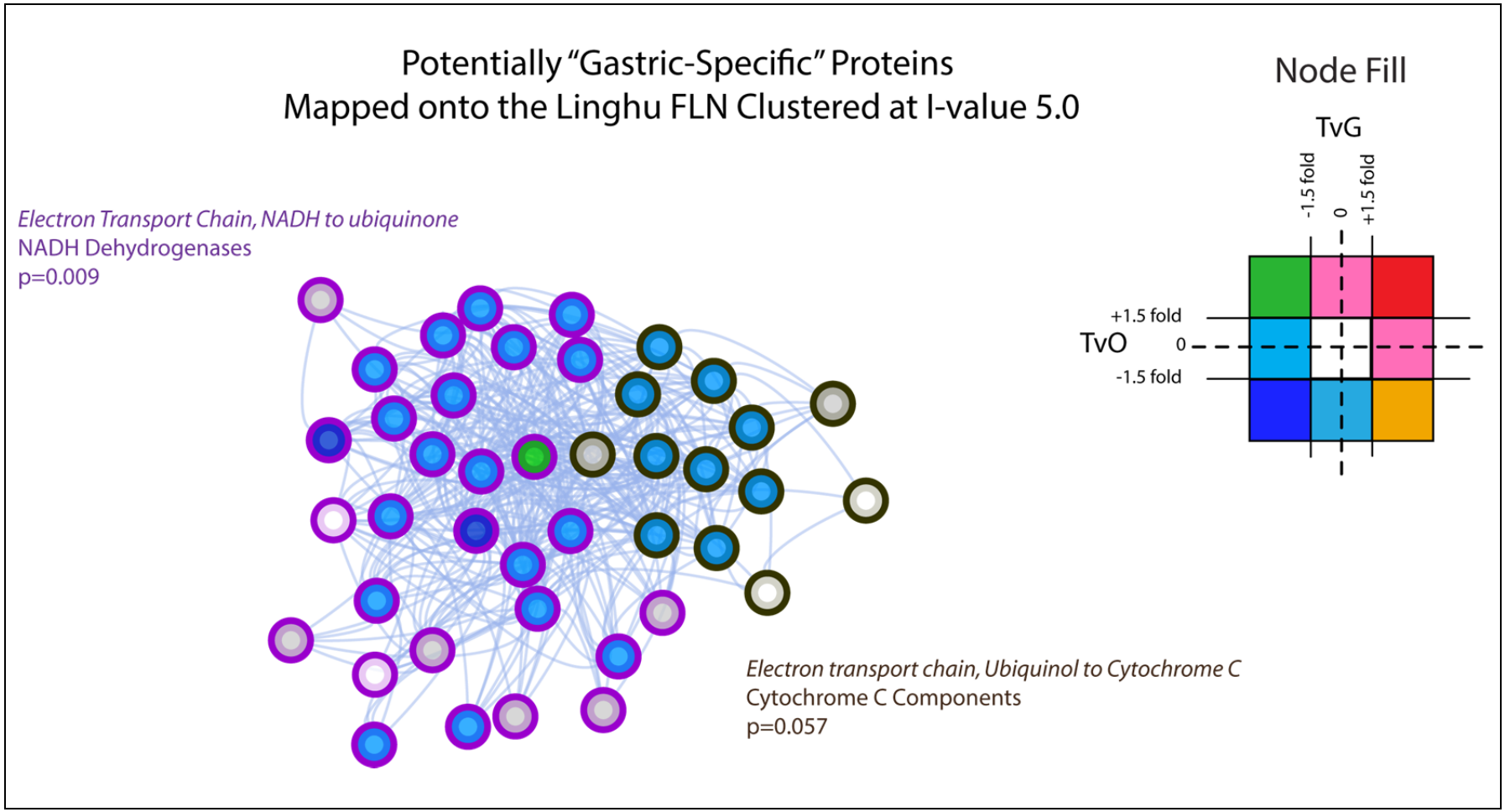

Figure 7-20 Significantly Enriched Functional Clusters in the "Gastric Specific" Genes. The top biological process ontology for each cluster is detailed in italics, and a summary of the cluster functional class and the FDR-corrected p-value for the enrichment of the cluster are displayed. Nodes have been filled according to TvO and TvG expression as in Figure 7-15 with node outlines corresponding to cluster membership.

One cluster identified in the 1.5 fold up-regulated network contained proteins associated with cell adhesion. This cluster contained both highly up and down-regulated proteins and was also identified using the 1.5-fold down-regulated gene list. The functional connectivity, relative expression and node identity are illustrated in Figure 7-21. The network has been manually scaled to reveal the connectivity and prevent node overlap. The network exhibits high connectivity, in keeping with a cluster containing members with closely related functions. Despite this, groups of proteins with predicted similar functions show divergent TvO and TvG expression ratios within the cluster. This suggests dynamic changes between cancer and normal tissues within these functional groups and a more complex regulatory structure. These dynamic changes are often visualised using a canonical pathway model incorporating feedback loops. Unfortunately the regulation of many of these genes is not well described, although there is some evidence for dysregulation of some of the cluster members in cancer[685,686]. This limits the current possibilities for a pathway-based visualisation incorporating the whole cluster and further studies are needed to clarify the roles of these nodes in OAC.

Similar to the previous network analysis, a striking feature of the clusters identified from the down-regulated gene lists is the widespread down-regulation of genes associated with oxidative phosphorylation and aerobic metabolism (Figure 7-17, Figure 7-18). As previously suggested, this may be evidence for the Warburg effect in this cancer[680,687]. Clusters associated with aerobic metabolism were also identified from the "gastric-specific" gene list (see discussion). Changes in the expression levels of metabolic enzymes may lead to cancer-specific vulnerabilities that may be therapeutic targets. To visualise the changes in metabolic enzyme expression in greater detail, all of the significantly dysregulated proteins were mapped to KEGG metabolic pathways (Figure 7-22, Figure 7-23 and Figure 7-24). This confirmed widespread downregulation of metabolic enzymes involved in aerobic respiration.

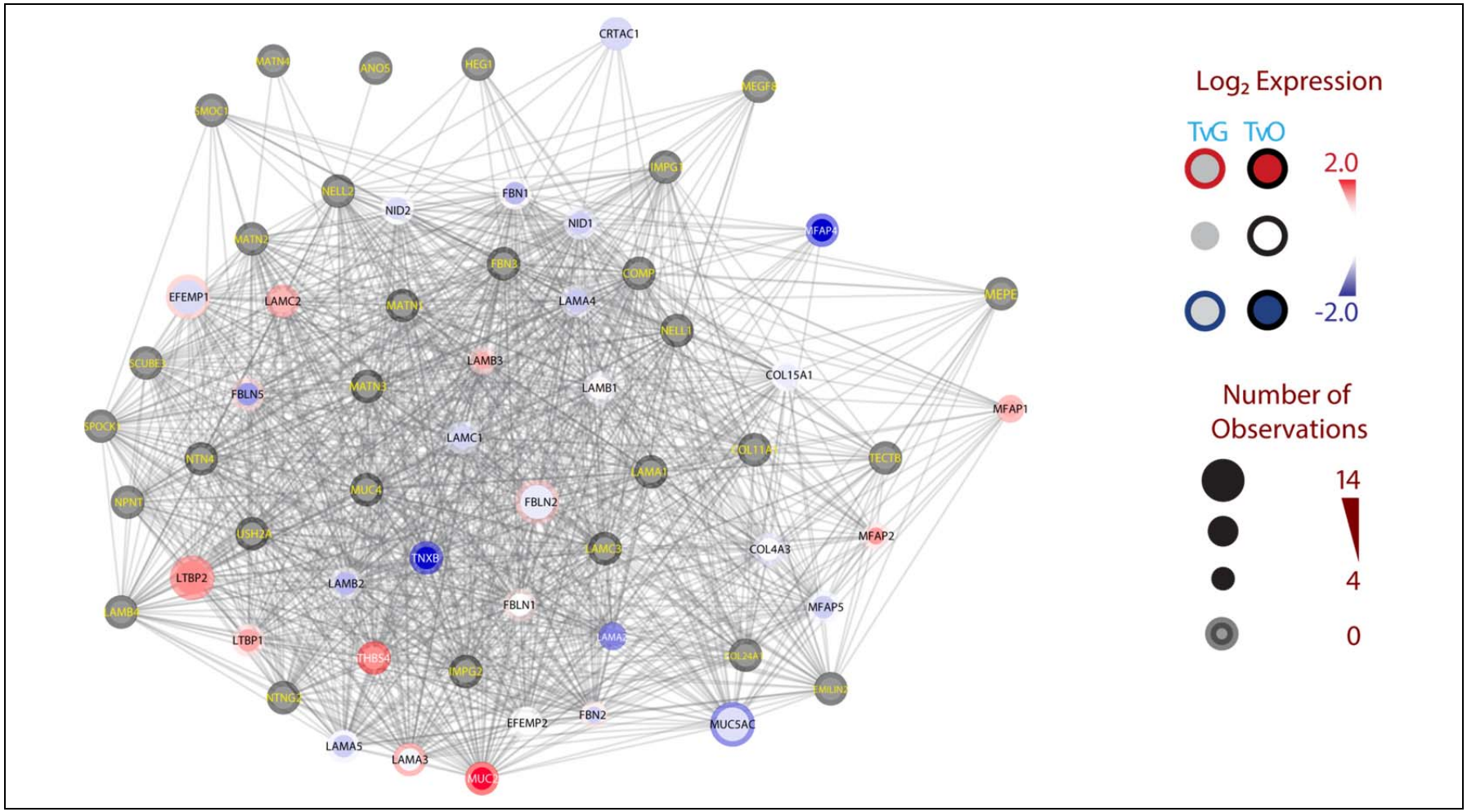

Figure 7-21 Relative Expression and Functional Connections of Genes in the Cell Adhesion Cluster (from Figure 7-15 and Figure 7-18). The figure legend demonstrates the corresponding Log_2 expression associated with the central (TvO) and outer (TvG) colour of each node. Proteins lacking expression data have been coloured grey and labelled with yellow text. Nodes have also been scaled in size according to the number of experimental observations and are labelled with gene names.

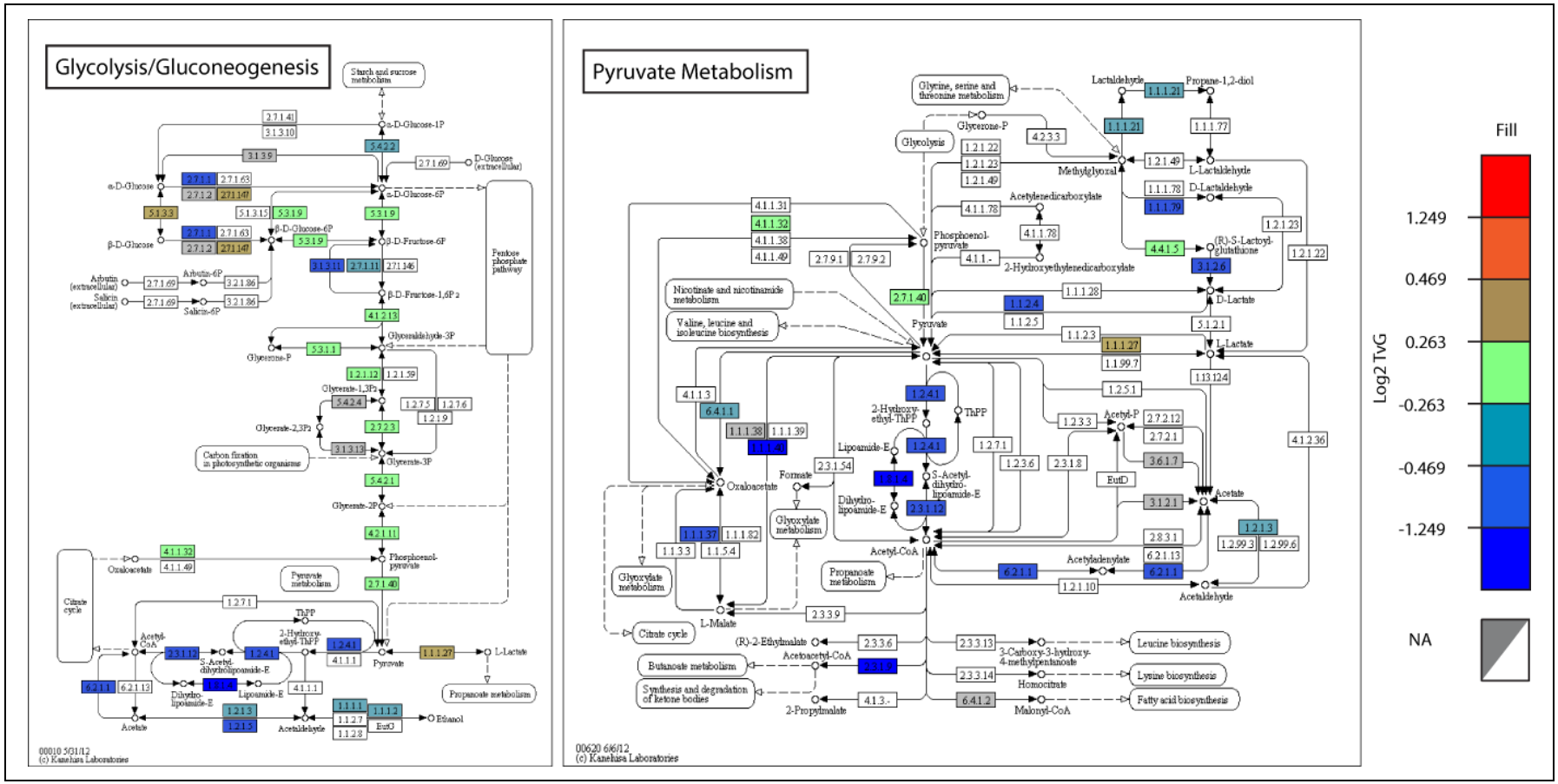

Figure 7-22 TvG Expression Mapped to KEGG Glycolytic and Pyruvate Metabolic Pathways. All significantly dysregulated proteins (TvO and TvG intra-experimental $p<0.05$) were mapped to KEGG identifiers and the associated TvG expression used to colour nodes in the KEGG metabolic pathways (mapping detailed in the legend). Similar patterns were observed with TvO expression mapping. Boxes represent metabolic enzymes labelled with KEGG identifiers, circles represent metabolites (annotated by name), and arrows represent reactions.

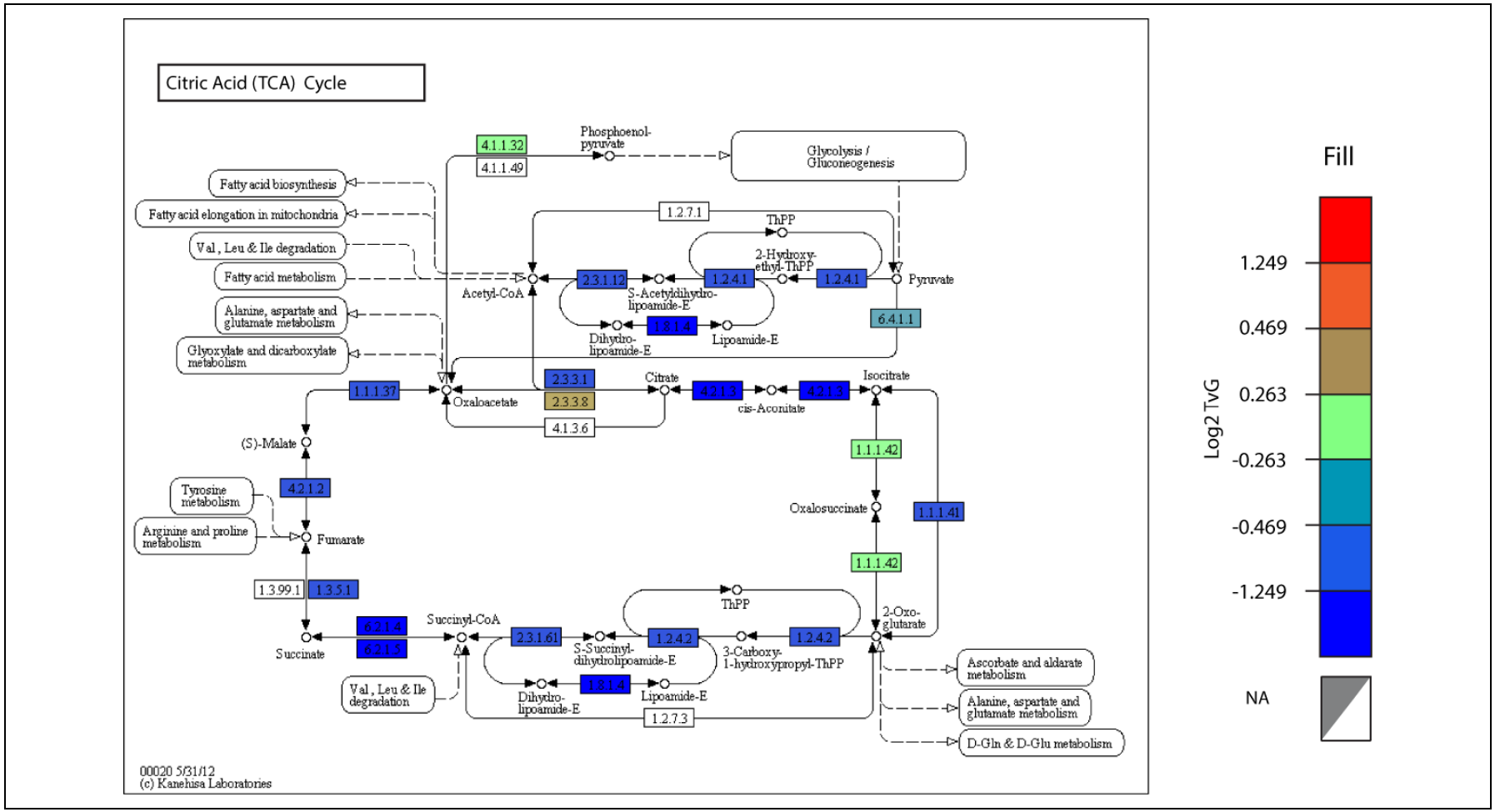

Figure 7-23 TvG Expression Mapped to the KEGG Citric Acid Cycle Pathway. All significantly dysregulated proteins (TvO and TvG intra-experimental p<0.05) were mapped to KEGG identifiers and the associated TvG expression used to colour nodes in the KEGG metabolic pathways (mapping detailed in the legend). Similar patterns were observed with TvO expression mapping. Boxes represent metabolic enzymes labelled with KEGG identifiers, circles represent metabolites (annotated by name), and arrows represent reactions.

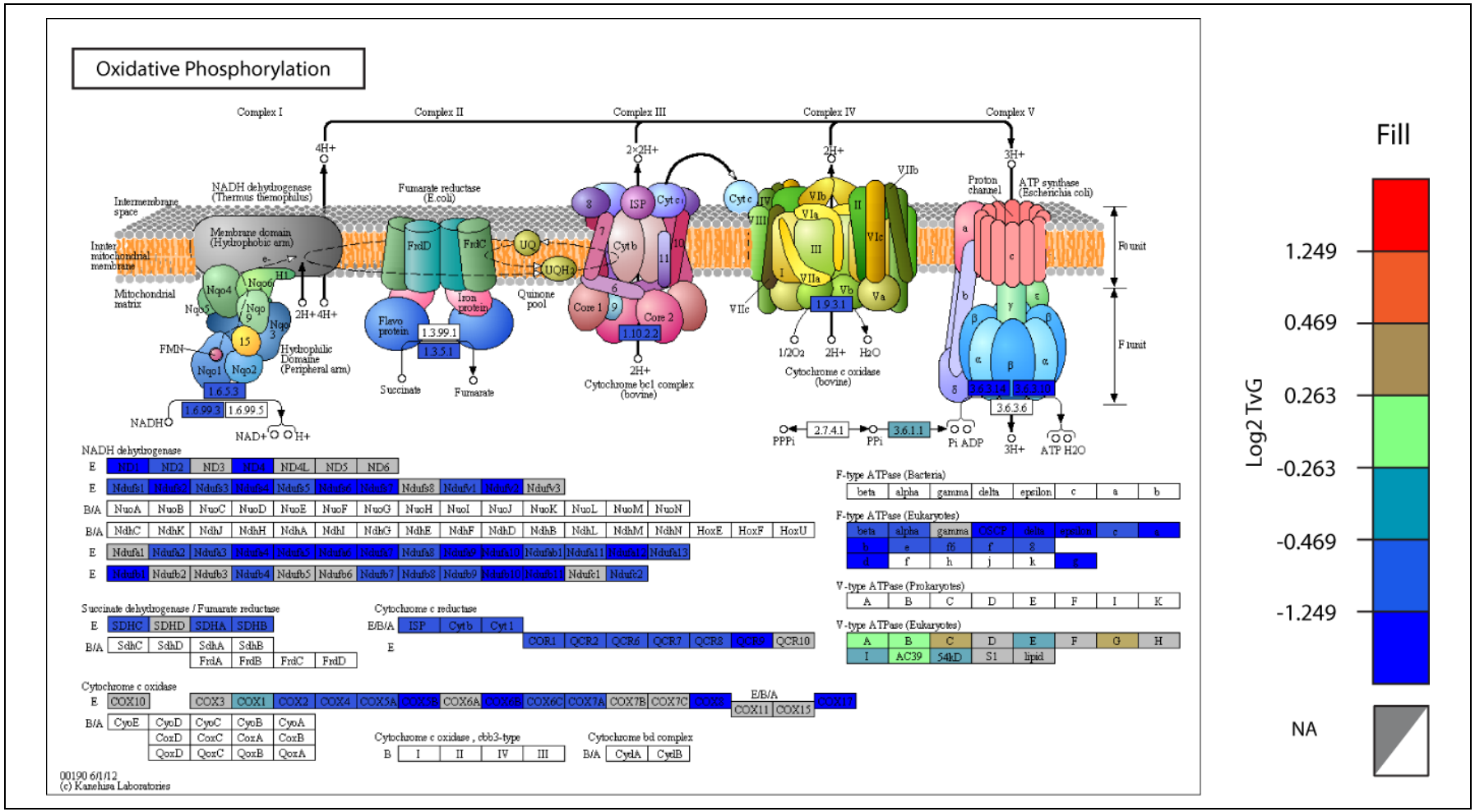

Figure 7-24 TvG Expression Mapped to the KEGG Oxidative Phosphorylation Pathway. All significantly dysregulated proteins (TvO and TvG intra-experimental p<0.05) were mapped to KEGG identifiers and the associated TvG expression used to colour nodes in the KEGG metabolic pathways (mapping detailed in the legend). Similar patterns were observed with TvO expression mapping. Boxes represent individual components of the metabolic enzyme complexes labelled with KEGG identifiers.

7.3.10 Generating *De Novo* Networks from Gene Lists

The previously applied network analysis method has used an existing functional linkage network to identify enriched clusters associated with biological functions. This method is well suited to incomplete datasets derived from proteomic studies as only a few cluster members are required to identify that cluster as enriched.

A further method to identify enriched functional pathways associated with dysregulated genes is to generate a network *de novo* from a dysregulated gene list using databases incorporating functional interaction data. This approach can employed using the STRING database[528]. Network connectivity can be enhanced by sequentially adding nodes with a high probability of interaction with the existing network members. Again this network method compensates for the incompleteness of a dataset. The STRING database also incorporates edge directionality if this data is available. Similar to the KEGG pathway analysis (Figure 7-22, Figure 7-23 and Figure 7-24) this can aid network interpretation although relative node expression cannot yet be visualised within STRING networks, limiting this feature.

The STRING database was used to construct a network from the 2 fold TvO and TvG up-regulated genes. When nodes with a very high probability of functional interaction were included (STRING score of ≥ 0.99)[688], a single highly connected network could be generated (Figure 7-25). This network suggests a central role for signalling via the epidermal growth factor family of receptor tyrosine kinases.

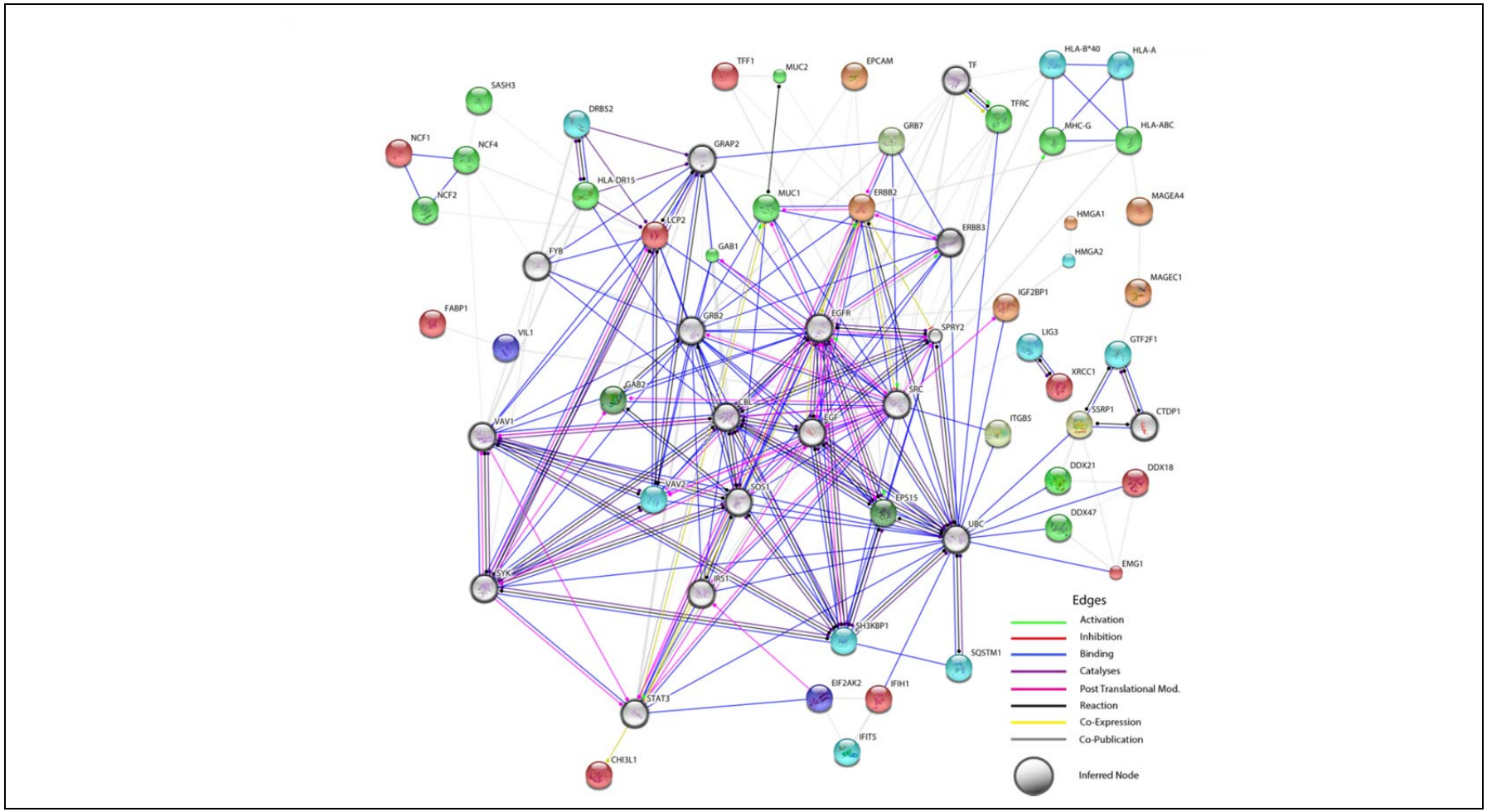

Figure 7-25 Inferred Network within the 2-fold TvO and TvG upregulated proteins with an intra-experimental FDR-p < 0.05. Edges and Inferred nodes were created using the STRING database. Unconnected and peripheral nodes have been removed. Nodes have been arbitrarily sized and coloured. Functional interactions have been summarised using various edges of different colours (see legend). A highly connected central network is apparent.

7.4 Discussion

This chapter has integrated previous work optimising tissue lysis and proteomic methods to analyse tissues from seven patients. The identification of over 6,000 proteins makes this the most extensive proteomic dataset in oesophageal cancer and emphasises the power of the mFASP TMT strategy. However, incomplete proteome coverage remains a significant limitation of shotgun proteomic studies and less than 1,000 proteins were identified in common between all samples. The technical reproducibility of protein quantitation was very high and a strategy was developed to identify significantly dysregulated proteins by comparison of TvO and TvG ratios with the corresponding technical replicates.

The combination of TvO and TvG ratios for each quantified protein allowed the generation of a 2D expression map. This map highlighted a group of potentially tumour-specific proteins (Figure 7-13). Although these have not been validated in this study, several have been previously shown to be over-expressed in OAC and represent potential therapeutic targets[689]. Indeed, immunotherapeutic trials are already underway with agents directed against EpCAM[690], GPA33[691,692] and MAGE[693] proteins in other cancer types. If the expression of these can be validated to be specific to OAC cells over surrounding tissues, there would be a compelling rationale to investigate these agents in OAC.

Examination of the 2D expression map suggested that genes with similar functions appeared in similar areas. This may be due to co-ordinated expression of functionally similar genes, e.g. neutrophil cytosol factors and melanoma antigens. This has been demonstrated previously in gene expression analysis and forms the basis of some methods used to infer gene function and visualise data[694,695]. It could be hypothesised that genes closely approximated on the map are likely to exhibit similar functions and this property could be used to infer function when it is not known. A network analysis was conducted to explore the biological processes associated with selected areas on the 2D expression map.

As previously discussed, an inflammatory infiltrate is commonly observed in OAC. Both the 1.5 and 2 fold upregulated lists from this analysis show enriched clusters

involved in the immune response. This supports the previous findings of enriched immune regulatory genes. Similarly, genes associated with RNA processing were enriched in both 1.5 fold and 2 fold upregulated lists in keeping with the previous network analysis. This may be secondary to an increased demand for functional protein associated with cell growth. In that case a concomitant enrichment in genes associated with protein translation and processing would be expected; although these clusters may be missed due to incomplete proteome coverage. Another possibility is that increased splicing of immature transcripts is present in OAC cells. Indeed one study reported over 50% of proteins demonstrated tissue-specific alternatively spliced isoforms when normal and cancer biopsies are compared[696]. These alternative-splice variants may be therapeutic targets or cancer-specific biomarkers in their own right as an independent proteomic study suggested[681].

The network analysis also confirmed widespread downregulation of genes associated with oxidative metabolism, suggestive of the Warburg effect. Higher levels of NADH dehydrogenase expression were noted in gastric epithelium compared to both tumour and normal oesophageal tissue. There may be gastric-specific metabolic changes responsible for this although this has not previously been described. Alternatively, the identified clusters may reflect a systemic difference in the tissue composition of biopsies. Muscle has been shown to express slightly higher levels of NADH dehydrogenases compared to gastrointestinal epithelium[697]. The stomach contains a thicker muscle layer than the oesophagus and if biopsies contained a higher proportion of muscle this may explain the relatively higher expression of theses enzymes. Both squamous oesophageal and tumour biopsies were taken tangentially to the muscle layer to prevent compromise of the circumferential resection margin. Therefore muscle was a relatively small component of the biopsies from these tissue types. Due to the pliability of the gastric mucosa in fresh resection specimens, biopsies were often taken perpendicular to the mucosal surface. An imbalance in the muscle component of biopsies could potentially explain the observed expression pattern for NADH dehydrogenases but would not account for the other metabolic clusters dysregulated between both tumour and normal oesophagus and tumour and normal gastric epithelium.

Gastric-specific epithelial markers including gastric intrinsic factor and mucin 5AC[697,698] also exhibited higher expression in gastric tissue compared to both OAC and normal oesophagus suggesting a significant proportion of each biopsy must have been epithelial. Frozen sections of each biopsy were reviewed by a consultant upper-GI pathologist prior to sample processing and no significant differences in relative epithelial/stromal composition were noted. The sections only demonstrate one plane through the biopsy, however and differences could still exist. To confirm the cellular origin of NADH dehydrogenase expression, IHC on patient matched normal gastric, oesophageal and OAC tissue could be performed.

Protein expression was visualised in the context of KEGG pathway maps. If it is assumed that metabolic enzyme expression correlates with reaction flux, the metabolism of glucose to pyruvate appeared unaffected between adenocarcinoma and normal gastric tissue (Figure 7-22). Similar appearances were observed between OAC and normal oesophageal tissue (not shown). The fate of pyruvate appeared to be significantly altered between the tissues with the conversion to lactic acid preferred in cancer cells over the generation of Acetyl-Coenzyme A (Acetyl-CoA). This would be expected to have significant consequences in shuttling glycolytic metabolites away from the citric acid cycle. Indeed this switch is further emphasised by the observation of a widespread suppression of citric acid cycle enzymes and electron transport chain components (Figure 7-22 and Figure 7-23). Interestingly the normally gluconeogenic reaction converting oxaloacetate to the pyruvate precursor phosphoenol pyruvate appears preserved. This generates a potential exit pathway for Acetyl-CoA generated from fatty acid metabolism or oxaloacetate generated from Aspartate or Asparagine.

These findings are in keeping with reports of rewired metabolic networks in cancer cells. The reasons for the shift in metabolism of pyruvate to lactic acid over oxidative phosphorylation are still a matter of debate. Convincing evidence has been provided for a survival advantage for tumour cells within their microenvironment produced by extracellular acidification[699]. The 18-fold lower yield of ATP from fermentation of glucose to lactic acid compared to complete oxidation to CO_2 via the citric acid cycle and electron transport chain is often mooted as the paradox of the

Warburg effect in a potentially resource-limited situation such as the tumour microenvironment. The resolution of this paradox has been proposed from substantially increased rate of ATP production by fermentation of glucose to lactic acid, resulting in a higher total yield per unit time than if oxidative phosphorylation is employed[700]. A diversion of glycolytic metabolites to generate nucleotides, amino acids and lipids to allow cell replication has also been proposed as a beneficial consequence[680].

The consequence of the metabolic changes in OAC cells also generates potential therapeutic targets. Glucose uptake is consistently elevated in OAC and forms the basis for the clinical staging investigation exploiting positron detection from the glucose analogue ^{18}Fluorodeoxyglucose[701]. Due to the hydrophilic nature of glucose, cellular uptake relies on transmembrane transporters. The GLUT1 glucose transporter has been published to be over-expressed in OAC compared to non-dysplastic Barrett's epithelium and surrounding normal tissue[702,703]. Consistent with this, GLUT1 was 1.2 fold over-expressed in OAC compared to normal squamous epithelium and 3.9 fold over-expressed in OAC compared to gastric epithelium. Both therapeutic antibodies and small molecule inhibitors of GLUT1-mediated glucose transport have been developed and demonstrate anti-cancer effects[704,705]. A significant challenge with GLUT1 based therapies is the off-target effect of suppression of glucose uptake by erythrocytes and neuronal cells although this toxicity was not reported in one recent study of a small molecule GLUT1 antagonist[706].

The KEGG pathway analysis revealed pyruvate metabolism to be a key difference between OAC and normal tissues (Figure 7-22 and Figure 7-23). The availability of glycolytic metabolites for biomass production can in part be governed by the rate of conversion from phosphoenolpyruvate to pyruvate by pyruvate kinase. This enzyme is active as a tetramer and two isoforms; PKM1 and PKM2, are generated by alternative splicing of primary transcripts from the same gene. Pyruvate kinase isoform detection was limited in this study by the requirement to detect peptides specifically derived from the variable exon resulting in a 40 amino acid segment unique to each isoform. A more specific, targeted proteomic method such as

selective reaction monitoring would be preferable for accurate quantification of tissue isoform expression[455]. Pyruvate kinase has been implicated in the metabolic switch between cancer and non-transformed cells; although the relative contribution of each isoform is still controversial[707,708]. While the PKM1 isoform is constitutively active, PKM2 can be allosterically controlled by variety of methods including direct binding by glycolytic metabolites[709]. This allosteric regulation of PKM2 has been shown to have a direct effect on cellular proliferation with the PKM2 activator, fructose 1,6-bisphosphate completely abrogating cell growth *in vitro*[710]. Activators of PKM2 therefore present compelling candidates as anti-cancer agents[711].

In this analysis, V-Type ATPases were not down-regulated along with other components of the electron transport chain (Figure 7-24). This may reflect their diverse role in transmembrane proton transport outside of mitochondrial oxidative phosphorylation and potential pro-oncogenic functions[712]. Reflecting this, a selective inhibitor of V-Type ATPases prevents invasion and metastasis and induces cancer cell apoptosis[713,714].

A *de novo* network was generated using the 2 fold TvO and TvG upregulated gene list and the STRING database. With the addition of high confidence inferred nodes, a network dominated by EGF-family receptor signalling was created. There are caveats associated with this finding, however, as random resamples of the identified non-dysregulated proteins were not compared with the dysregulated proteins to confirm the specificity of the association with EGF-family genes. This approach, employed in the previous network analysis (Chapter 6), would control for the potential bias towards identifying highly studied proteins such as receptor tyrosine kinases.

Nevertheless, several previous studies have implicated EGFR signalling in OAC. Using tissue microarrays, EGFR has been validated by IHC along with two other genes to aid in the determination of long term prognosis after resection of OAC[715]. EGFR has been demonstrated to be overexpressed in 15-50% of patients with OAC with amplification in 20%[83,716]. Similarly ErbB2 is amplified in around 20% of patients with oesophageal or OGJ adenocarcinoma[717]. The monoclonal antibodies

designed to inhibit both EGFR and ErbB2; Lapatinib, or EGFR alone; Erlotinib, induce cell death in OAC cell lines showing high levels of target expression[718].

Despite the strong preclinical rationale for EGFR inhibition in OAC, a phase III clinical trial (REAL-3) of the monoclonal anti-EGFR antibody panitumumab recently reported a poorer prognosis with the addition of the EGFR inhibitor to the standard regimen or Epirubicin, Oxaliplatin and Capecitabine (EOX)[719]. This may have been secondary to a synergistic effect of panitumumab on toxicity from EOX or the lack of patient stratification by EGFR expression.

Resistance to EGFR inhibition despite EGFR amplification has also been demonstrated in KRAS mutant cancers[720]. This is unlikely to explain the data in oesophageal cancer as KRAS mutations are generally infrequent in this cancer[218] and in the subgroup of around 1/3 of patients assessed for KRAS allele genotype in the REAL-3 trail, mutations were detected in less than 6% of tumours[719].

The failure of this therapy in patients underscores the complexity of cellular signalling and the potential for secondary compensation when agents directed against a single receptor tyrosine kinase (RTK) are used. Even in tumours with apparent clinical response to therapies targeted to RTKs, the later emergence of resistance is common and in some cases universal[160,361,721,722]. This is likely due to selective pressure leading to the expansion of resistant clones, present initially in small numbers.

A novel strategy embracing the complexity of cell signalling at the network level identified time-dependent changes in the response to EGFR inhibition that offered a window of opportunity for effective sequential therapy. When applied to triple-negative breast cancer cells, pre-treatment by inhibition of EGFR, either using a small molecule or siRNA, conditioned cells to a network state sensitive to the DNA damaging agent doxorubicin[723]. Importantly co-treatment did not produce the same effect. Similar approaches may be more successful in OAC than current strategies.

Chapter 8: Conclusions and Future Work

Oesophageal adenocarcinoma remains a highly lethal disease with a 5 year survival rate of less than 15%[1,127]. The early dissemination of OAC mandates that effective systemic therapies are required to improve survival. The majority of OACs have not been demonstrated to express a single dominant oncogene or kinase limiting the application of existing targeted therapies. This thesis has therefore focused on strategies to identify therapeutic targets in this cancer.

Mutation of TP53 has been demonstrated in over 75% of OACs generating a possibility for tumour-specific therapy if a synthetic lethal interactor can be identified[550]. Published *in vitro* work proposed PLK-1 as a candidate[220].

8.1 Mutation of TP53 Suggests a Synthetic Lethal Strategy for OAC

PLK-1 was confirmed to be over-expressed in verified OAC cells and tissue compared to primary squamous cells, non-transformed Barrett's epithelial cells and Barrett's and normal squamous oesophageal tissue. Mutation of p53 was associated with over-expression of PLK-1 in OAC and ovarian cancer providing *in vivo* evidence of a possible synthetic lethal therapeutic strategy through targeting PLK-1 (Chapter 3).

The identification of cancer cell cytotoxicity compared to cytostasis or reduction in proliferation was deemed important to evaluate potential therapeutic efficacy. The AlamarBlue viability assay was adapted for this purpose and was demonstrated to be sensitive, accurate and specific.

Inhibition of PLK-1 expression by siRNA reduced viability in oesophageal cell lines with the magnitude of reduction related to the degree of PLK-1 suppression. An established (BI2536) and a novel, short half-life, PLK-1 inhibitor (PLK-1A) reduced proliferation in short-term viability assays with constant drug exposure. An enhanced effect was observed in oesophageal cancer cell lines compared to primary squamous cells and this relative selectivity appeared marginally greater with PLK-1A.

Assays using constant drug concentrations *in vitro* do not mimic *in vivo* cancer cell drug exposure and therefore a variety of intermittent exposure experiments were conducted. Sensitivity to PLK-1A was enhanced by pulsed administration and this effect appeared to be specific to PLK-1A compared to BI2536. The relative cancer cell specificity and the enhanced therapeutic effect with pulsed administration for PLK-1A provide a clear rationale for further development towards *in vivo* human studies. Augmenting the therapeutic window by pulsed administration of a short half-life compound is a new strategy for PLK-1 inhibitors and if verified with other compounds may help to limit the toxicity which has plagued Phase I and II trials of BI2536 and other similar long half-life inhibitors. The *in vitro* methods adopted for the assay of proliferation and cytotoxicity in this study and the use of a panel of both non-transformed and transformed oesophageal cells also provide a framework for the interrogation of other future candidate therapeutic targets.

Evaluation of pulsed administration regimes *in vivo* would help to confirm these cell line derived data. For oesophageal cancer this would likely rely on the use of nude mouse xenografts. Tumour size measurements could be used to confirm efficacy, however, PLK-1 inhibitors are associated with significant hematological toxicity in human trials and immuno-competent models would be required to capture this and allow iteration of dosing schedules. The recently described transgenic mouse model of Barrett's carcinogenesis could, in contrast, be used to optimise the therapeutic index for PLK-1 inhibitors and other candidates in preclinical trials[102].

From this work it appears that oesophageal cancer offers a disease indication for this class of inhibitors. This finding has driven the development of an oesophageal cancer focused Academic/Industry collaborative programme "First-In-Human directed pre-clinical development of anti-cancer agent CYC140". This was recently supported by Biomedical Catalyst Funding as an Early Stage Award from the MRC and the Technology Strategy Board. This project will focus on further characterisation of *in vitro* resistance mechanisms, optimisation of drug dosing regimens including combination therapy and evaluation of large mammal toxicity with the PLK-1A drug and more potent derivatives. It is hoped that this will result in the progression of this class of compounds into clinical trials in oesophageal cancer.

8.2 Using Quantitative Shotgun Proteomics to Identify OAC-specific Proteins

Although highly expressed in OAC compared to surrounding tissues, PLK-1 is still expressed in non-transformed cells due to its central role in the cell cycle. The highly conserved structure of the kinase domain also creates difficulties in generating truly kinase-specific inhibitors. Both of these limitations increase the probability of off-target effects and toxicity.

An alternative strategy for therapeutic target identification is comparative proteomic profiling to identify proteins specifically expressed in OAC cells. These could form the basis for the development of targeted immunotherapeutics with the potential for high specificity and low toxicity.

At the commencement of this project, no studies had been published using entirely shotgun proteomic methods on OAC tissue. An oesophageal tissue resource was established and the feasibility and accuracy of a quantitative shotgun proteomic strategy using patient-matched OAC and normal squamous tissue assessed (Chapter 4). A relative quantitative approach was employed and many proteins previously published as dysregulated in OAC were identified, verifying the validity and potential power of this approach. The quantitative accuracy was also confirmed from the expression pattern of six proteins (AGR2, GRP78, SAMHD1, D4GDI, HSP27 and TGM3) determined by IHC using validated antibodies on the same tissues used for proteomic analysis. The IHC expression pattern revealed insights into the cellular origin of dysregulated proteins and underscores the importance of IHC for validation of proteomic and transcriptomic studies of human tissue.

The proteomic method was subsequently optimised to increase the number of protein identifications and improve the sensitivity to detect OAC-specific proteins. The combination of the mFASP method and an increased number of OFF-GEL fractions significantly enhanced proteome coverage (Chapter 5). A published label-free systematic sequencing method, PAcIFIC, had been proposed to increase protein identifications and dynamic range but in this investigation, the isobaric quantitative

strategy with sample fractionation by isoelectric point appeared to be superior to PAcIFIC (Chapter 5).

Label-free proteomic methods are becoming increasingly common, however, and have the advantage of minimal sample preparation and applicability to low yield sample types. In the label-free approach, each sample is analysed individually. A fundamental limitation of proteomic studies on human tissues is the identification of only a subset of all proteins expressed in a sample. This leads to the central problem of identifying which proteins have been missed but are present in the sample and which are truly not expressed and potentially tumour specific.

To address this, a Bayesian statistical approach was developed in collaboration with Dr. Ian Overton (IGMM, Edinburgh University). This allowed the prediction of a threshold of tumour protein abundance associated with a high probability of tumour-specific expression when the protein was not detected in normal oesophageal tissue and vice versa (Chapter 5). Several of the proposed candidate tumour specific proteins were independently demonstrated to be highly expressed in OAC compared to surrounding normal tissues (Chapter 4, 7) lending support for the validity of the method. This approach was not, however, formally externally validated. This could be performed by obtaining serial PAcIFIC proteomes from a sample of known protein composition. A challenge would be in creating the necessary sample of known composition with several thousand different proteins.

A list of all identified proteins with their relative tissue expression was generated from the PAcIFIC data using the Bayesian approach. The quantitative methods of spectral counting and TMT reporter ion intensity were not comparable, potentially due to limitations in the spectral counting method. The PAcIFIC and TMT data were therefore subsequently treated separately.

A third strategy to identify therapeutic targets in OAC is to identify dysregulated targetable pathways specific to OAC tissue. Using quantitative proteomic data has the advantage of potentially greater functional relevance but incomplete proteomic

datasets make the identification of dysregulated pathways a challenge. This was addressed using network analysis methods.

8.3 Identifying Dysregulated Biological Processes in OAC using Quantitative Proteomic Data

Strategies originally developed for gene-set and gene-network enrichment analysis were adapted to identify functional pathways from incomplete proteomic data. A genome-scale functional linkage network was clustered to generate biological functional clusters – verified by the significant enrichment of DAVID biological process ontologies.

Multiple dysregulated functional clusters were identified by network analysis of the mFASP TMT and PAcIFIC proteomic data. The identified clusters were highly enriched for biological processes including inflammatory signalling and oxidative phosphorylation which may represent evidence of targetable pathways. The findings from the Linghu functional linkage network were replicated using an independent human network, providing support for their validity. Although apparently robust, these data were obtained from only seven patients samples and can only be considered preliminary until validated in tissue from a larger cohort of patients with OAC.

The PAcIFIC and pilot TMT data revealed significant variability in protein identification between patient's samples. High technical reproducibility from the TMT data suggested a large component of this was biological variation. The optimised (mFASP) TMT method was therefore used on samples from a further five patients and all mFASP TMT data considered together (Chapter 7).

This study of 7 patients' tissues identified and quantified 6,339 unique protein groups, the most comprehensive proteomic study to date in oesophageal cancer. The mFASP TMT method demonstrated high technical reproducibility but incomplete proteome coverage was still a factor. Despite this, tumour-selective expression was proposed for a subset of proteins containing both novel and existing immunotherapeutic candidates.

One of these, EpCAM, has been previously demonstrated to be highly expressed in OAC tissue compared to normal squamous and gastric epithelium[724,725]. EpCAM-directed immunotherapeutics have been licensed for the treatment of malignant ascites and are under investigation for several other malignancies[726]. If EpCAM can be validated as specific for tumour cells in OAC, existing EpCAM therapies could enter clinical trials in this disease.

A subsequent network analysis of the complete mFASP TMT dataset confirmed the prominent role of inflammatory signalling and RNA processing in these tumours and widespread changes in metabolic pathways consistent with the Warburg effect were identified. This study illustrates the potential of network and proteomic methods to uncover dysregulated signalling pathways and candidate therapeutic targets.

In conclusion, this thesis examines three strategies to identify therapeutic targets for OAC and reveals several novel findings. The evaluation of PLK-1 as a therapeutic target in OAC will continue as part of an Academic-Industry collaborative programme. Using comparative proteomics, several compelling immunotherapeutic candidates are proposed and require to be validated in a larger cohort of patient's samples. To achieve this, a tissue-microarray comprising of FFPE cores from OAC and patient-matched normal and metastatic tissues is being constructed to validate the relative specificity of expression of candidate targets by IHC. Candidates will be taken forward for *in vitro* study using the approaches employed in the evaluation of PLK-1 as a therapeutic target. Finally, network analysis of proteomic data proposed deranged glucose metabolism in OAC through changes consistent with the Warburg effect. The role of PKM2 as a key mediator of this switch and potential target is being actively explored[710]. Together these strategies reveal new areas for therapeutic innovation. It is hoped they will inspire future efforts to develop effective therapeutics that will ultimately improve the survival from OAC.

Chapter 9: References

1. Information Statistics Division. *Cancer in Scotland*, (NHS National Services Scotland, Edinburgh, 2010).
2. Jemal, A., *et al.* Global cancer statistics. *CA Cancer J Clin* **61**, 69-90 (2011).
3. Pohl, H. & Welch, H. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst* **97**, 142-146 (2005).
4. Lepage, C., Rachet, B., Jooste, V., Faivre, J. & Coleman, M.P. Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *Am J Gastroenterol* **103**, 2694-2699 (2008).
5. Crane, L.M., *et al.* Oesophageal cancer in The Netherlands: increasing incidence and mortality but improving survival. *Eur J Cancer* **43**, 1445-1451 (2007).
6. Bosetti, C., *et al.* Trends in oesophageal cancer incidence and mortality in Europe. *Int J Cancer* **122**, 1118-1129 (2008).
7. Stavrou, E.P., McElroy, H.J., Baker, D.F., Smith, G. & Bishop, J.F. Adenocarcinoma of the oesophagus: incidence and survival rates in New South Wales, 1972-2005. *Med J Aust* **191**, 310-314 (2009).
8. Islami, F., *et al.* Tea drinking habits and oesophageal cancer in a high risk area in northern Iran: population based case-control study. *BMJ* **338**, b929 (2009).
9. Westlake, S. Cancer incidence and mortality in the United Kingdom and constituent countries, 2004-06. *Health Stat Q*, 56-62 (2009).
10. Rutegård, M., Nordenstedt, H., Lu, Y., Lagergren, J. & Lagergren, P. Sex-specific exposure prevalence of established risk factors for oesophageal adenocarcinoma. *Br J Cancer* **103**, 735-740 (2010).
11. Chandanos, E. & Lagergren, J. The mystery of male dominance in oesophageal cancer and the potential protective role of oestrogen. *Eur J Cancer* **45**, 3149-3155 (2009).
12. http://www.cancerresearchuk.org/cancerinfo/cancerstats/types/oesophagus/incidence/#source5
13. Sakata, K., *et al.* Smoking, alcohol drinking and esophageal cancer: findings from the JACC Study. *J Epidemiol* **15 Suppl 2**, S212-219 (2005).
14. Zhang, H.Z., Jin, G.F. & Shen, H.B. Epidemiologic differences in esophageal cancer between Asian and Western populations. *Chin J Cancer* **31**, 281-286 (2012).
15. Guo, F., *et al.* Human papillomavirus infection and esophageal squamous cell carcinoma: a case-control study. *Cancer Epidemiol Biomarkers Prev* **21**, 780-785 (2012).
16. Syrjänen, K.J. HPV infections and oesophageal cancer. *J Clin Pathol* **55**, 721-728 (2002).
17. Syrjänen, K. & Syrjänen, S. Detection of human papillomavirus in esophageal papillomas: systematic review and meta-analysis. *APMIS* **121**, 363-374 (2013).
18. Lee, C.H., *et al.* Carcinogenetic impact of alcohol intake on squamous cell carcinoma risk of the oesophagus in relation to tobacco smoking. *Eur J Cancer* **43**, 1188-1199 (2007).
19. Cui, R., *et al.* Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology* **137**, 1768-1775 (2009).
20. Wu, C., *et al.* Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat Genet* **44**, 1090-1097 (2012).
21. Blaydon, D.C., *et al.* RHBDF2 mutations are associated with tylosis, a familial esophageal cancer syndrome. *Am J Hum Genet* **90**, 340-346 (2012).
22. Iwaya, T., Maesawa, C., Ogasawara, S. & Tamura, G. Tylosis esophageal cancer locus on chromosome 17q25.1 is commonly deleted in sporadic human esophageal cancer. *Gastroenterology* **114**, 1206-1210 (1998).
23. Lagergren, J., Bergström, R., Lindgren, A. & Nyrén, O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med* **340**, 825-831 (1999).
24. Anderson, L.A., *et al.* Risk factors for Barrett's oesophagus and oesophageal adenocarcinoma: results from the FINBAR study. *World J Gastroenterol* **13**, 1585-1594 (2007).

25. Kubo, A. & Corley, D.A. Body mass index and adenocarcinomas of the esophagus or gastric cardia: a systematic review and meta-analysis. *Cancer Epidemiol Biomarkers Prev* **15**, 872-878 (2006).
26. Fisher, B.L., Pennathur, A., Mutnick, J.L. & Little, A.G. Obesity correlates with gastroesophageal reflux. *Dig Dis Sci* **44**, 2290-2294 (1999).
27. Ryan, A.M., *et al.* Barrett esophagus: prevalence of central adiposity, metabolic syndrome, and a proinflammatory state. *Ann Surg* **247**, 909-915 (2008).
28. Donohoe, C.L., Pidgeon, G.P., Lysaght, J. & Reynolds, J.V. Obesity and gastrointestinal cancer. *Br J Surg* **97**, 628-642 (2010).
29. Björntorp, P. The regulation of adipose tissue distribution in humans. *Int J Obes Relat Metab Disord* **20**, 291-302 (1996).
30. Engel, L.S., *et al.* Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst* **95**, 1404-1413 (2003).
31. Hoeijmakers, J.H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366-374 (2001).
32. DeMeester, S. & DeMeester, T. Columnar mucosa and intestinal metaplasia of the esophagus: fifty years of controversy. *Ann Surg* **231**, 303-321 (2000).
33. DeMeester, T. Clinical biology of the Barrett's metaplasia, dysplasia to carcinoma sequence. *Surg Oncol* **10**, 91-102 (2001).
34. Goldman, A., *et al.* A novel mechanism of acid and bile acid-induced DNA damage involving Na+/H+ exchanger: implication for Barrett's oesophagus. *Gut* **59**, 1606-1616 (2010).
35. Souza, R.F., Krishnan, K. & Spechler, S.J. Acid, bile, and CDX: the ABCs of making Barrett's metaplasia. *Am J Physiol Gastrointest Liver Physiol* **295**, G211-218 (2008).
36. Spechler, S.J. & Goyal, R.K. Barrett's esophagus. *N Engl J Med* **315**, 362-371 (1986).
37. Murphy, S.J., *et al.* Have patients with esophagitis got an increased risk of adenocarcinoma? Results from a population-based study. *World J Gastroenterol* **11**, 7290-7295 (2005).
38. Hawe, A., Payne, W.S. & LH, W. Adenocarcinoma in the columnar epithelial lined lower (Barrett) esophagus. *Thorax* **28**, 511-514 (1973).
39. Tileston, W. Peptic ulcer of the oesophagus. *Am J Med Sci* **132**, 240–265 (1906).
40. Barrett, N.R. The lower esophagus lined by columnar epithelium. *Surgery* **41**, 881-894 (1957).
41. Paull, A., Trier, J.S. & Dalton, M.D. The histologic spectrum of Barrett's esophagus. *N Engl J Med* **295**, 476–480 (1976).
42. Reid, B.J. Barrett's esophagus and esophageal adenocarcinoma. *Gastroenterol Clin North Am* **20**, 817–834 (1991).
43. Bhat, S., *et al.* Risk of malignant progression in Barrett's esophagus patients: results from a large population-based study. *J Natl Cancer Inst* **103**, 1049-1057 (2011).
44. Weinstein, W.M. & Ippoliti, A.F. The diagnosis of Barrett's esophagus: goblets, goblets, goblets. *Gastrointest Endosc* **44**, 91-95 (1996).
45. Takubo, K., Aida, J. & Naomoto, Y. Cardiac rather than intestinal-type background in endoscopic resection specimens of minute Barrett adenocarcinoma. *Hum Pathol* **40**, 65-74 (2009).
46. Liu, W., Hahn, H. & Odze, R.D. Metaplastic esophageal columnar epithelium without goblet cells shows DNA content abnormalities similar to goblet cell-containing epithelium. *Am J Gastroenterol* **104**, 816–824 (2009).
47. Gatenby, P.A., Ramus, J.R., Caygill, C.P., Shepherd, N.A. & Watson, A. Relevance of the detection of intestinal metaplasia in non-dysplastic columnar-lined oesophagus. *Scand J Gastroenterol* **43**, 524-530 (2008).
48. Playford, R.J. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett's oesophagus. *Gut* **55**, 442 (2006).
49. Spechler, S.J., *et al.* American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* **140**, 1084-1091 (2011).
50. Haggitt, R.C. Barrett's Esophagus, dysplasia, and adenocarcinoma. *Human Pathology* **25**, 982-993 (1994).

51. Hvid-Jensen, F., Pedersen, L., Drewes, A.M., Sørensen, H.T. & Funch-Jensen, P. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med* **365**, 1375-1383 (2011).
52. Desai, T.K., *et al.* The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: a meta-analysis. *Gut* **61**, 970-976 (2012).
53. Wani, S., *et al.* Risk factors for progression of low-grade dysplasia in patients with Barrett's esophagus. *Gastroenterology* **141**, 1179-1186, 1186.e1171 (2011).
54. Pennathur, A., Landreneau, R.J. & Luketich, J.D. Surgical aspects of the patient with high-grade dysplasia. *Semin Thorac Cardiovasc Surg* **17**, 326-332 (2005).
55. Reid, B.J., *et al.* Endoscopic biopsy can detect high-grade dysplasia or early adenocarcinoma in Barrett's esophagus without grossly recognizable neoplastic lesions. *Gastroenterology* **94**, 81-90 (1988).
56. Schnell, T.G., *et al.* Long-term nonsurgical management of Barrett's esophagus with high-grade dysplasia. *Gastroenterology* **120**, 1607-1619 (2001).
57. Konda, V.J., *et al.* Is the risk of concomitant invasive esophageal cancer in high-grade dysplasia in Barrett's esophagus overestimated? *Clin Gastroenterol Hepatol* **6**, 159-164 (2008).
58. Tharavej, C., *et al.* Predictive factors of coexisting cancer in Barrett's high-grade dysplasia. *Surg Endosc* **20**, 439-443 (2006).
59. Dahlberg, P.S., *et al.* Gene Expression Profiles in Esophageal Adenocarcinoma. *Annals of Thoracic Surgery* **77**, 1008-1015 (2004).
60. Kimchi, E.T., *et al.* Progression of Barrett's Metaplasia to Adenocarcinoma Is Associated with the Suppression of the Transcriptional Programs of Epidermal Differentiation. *Cancer Research* **65**, 3146-3154 (2005).
61. Wang, S., *et al.* Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis. *Oncogene* **25**, 3346-3356 (2006).
62. van den Akker, E., *et al.* Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* **129**, 2181-2193 (2002).
63. Silberg, D.G., *et al.* Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. *Gastroenterology* **122**, 689-696 (2002).
64. Stairs, D.B., *et al.* Cdx1 and c-Myc foster the initiation of transdifferentiation of the normal esophageal squamous epithelium toward Barrett's esophagus. *PLoS One* **3**, e3534 (2008).
65. Zhang, H.Y., Spechler, S.J. & Souza, R.F. Esophageal adenocarcinoma arising in Barrett esophagus. *Cancer Lett* **275**, 170-177 (2009).
66. di Pietro, M. & Fitzgerald, R.C. Barrett's oesophagus: an ideal model to study cancer genetics. *Hum Genet* **126**, 233-246 (2009).
67. Koppert, L.B., Wijnhoven, B.P., van Dekken, H., Tilanus, H.W. & Dinjens, W.N. The molecular biology of esophageal adenocarcinoma. *J Surg Oncol* **92**, 169-190 (2005).
68. Maley, C.C., *et al.* Selectively advantageous mutations and hitchhikers in neoplasms: p16 mutations are selected in Barrett's Esophagus. *Cancer Research* **64**, 3414-3427 (2004).
69. Schulmann, K., *et al.* Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. *Oncogene* **24**, 4138-4148 (2005).
70. Bian, Y.S., Osterheld, M.C., Fontolliet, C., Bosman, F.T. & Benhattar, J. p16 inactivation by methylation of the CDKN2A promoter occurs early during neoplastic progression in Barrett's esophagus. *Gastroenterology* **122**, 1113-1121 (2002).
71. Reid, B.J., *et al.* Predictors of progression in Barrett's esophagus II: baseline 17p (p53) loss of heterozygosity identifies a patient subset at increased risk for neoplastic progression. *American Journal of Gastroenterology* **96**, 2839-2848 (2001).
72. Wong, D.J., *et al.* p16 (INK4a) lesions are common early abnormalities that undergo clonal expansion in Barrett's metaplastic epithelium. *Cancer Research* **61**, 8284-8289 (2001).
73. Leedham, S., *et al.* Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut* **57**, 1041-1048 (2008).

74. Reid, B.J., *et al.* Flow-cytometric and histological progression to malignancy in Barrett's esophagus: prospective endoscopic surveillance of a cohort. *Gastroenterology* **102**, 1212-1219 (1992).
75. Sikkema, M., *et al.* Aneuploidy and overexpression of Ki67 and p53 as markers for neoplastic progression in Barrett's esophagus: a case-control study. *Am J Gastroenterol* **104**, 2673-2680 (2009).
76. Bird-Lieberman, E.L., *et al.* Population-based study reveals new risk-stratification biomarker panel for Barrett's esophagus. *Gastroenterology* **143**, 927-935.e923 (2012).
77. Arber, N., *et al.* Increased expression of the cyclin D1 gene in Barrett's esophagus. *Cancer Epidemiology Biomarkers and Prevention* **5**, 457-459 (1996).
78. Geddert, H., Heep, H.J., Gabbert, H.E. & Sarbia, M. Expression of cyclin B1 in the metaplasia–dysplasia–carcinoma sequence of Barrett's esophagus. *Cancer* **94**, 212-218 (2002).
79. Lao-Sirieix, P., Lovat, L. & Fitzgerald, R.C. Cyclin A immunocytology as a risk stratification tool for Barrett's esophagus surveillance. *Clinical Cancer Research* **13**, 659-665 (2007).
80. Sarbia, M., *et al.* Expression of cyclin E in dysplasia, carcinoma, and nonmalignant lesions in Barrett's esophagus. *Cancer* **86**, 2597-2601 (1999).
81. Sarbia, M., Tekin, U., Zeriouh, M. & Gabbert, H.E. Expression of the RB protein, allelic imbalance of the RB gene and amplification of the CDK4 gene in metaplasias, dysplasias and carcinomas in Barrett's oesophagus. *Anticancer Research* **21**, 387-392 (2001).
82. Goh, X.Y., *et al.* Integrative analysis of array-comparative genomic hybridisation and matched gene expression profiling data reveals novel genes with prognostic significance in oesophageal adenocarcinoma. *Gut* **60**, 1317-1326 (2011).
83. Paterson, A.L., *et al.* A systematic approach to therapeutic target selection in oesophago-gastric cancer. *Gut* **62**, 1415-1424 (2013).
84. Rees, J.R., Onwuegbusi, B.A., Save, V.E., Alderson, D. & Fitzgerald, R.C. In vivo and in vitro evidence for transforming growth factor-beta1-mediated epithelial to mesenchymal transition in esophageal adenocarcinoma. *Cancer Research* **66**, 9583-9590 (2006).
85. Boonstra, J.J., *et al.* Verification and unmasking of widely used human esophageal adenocarcinoma cell lines. *J Natl Cancer Inst* **102**, 271-274 (2010).
86. Boonstra, J.J., *et al.* Mistaken identity of widely used esophageal adenocarcinoma cell line TE-7. *Cancer Research* **67**, 7996-8001 (2007).
87. Lynam-Lennon, N., *et al.* Alterations in DNA repair efficiency are involved in the radioresistance of esophageal adenocarcinoma. *Radiat Res* **174**, 703-711 (2010).
88. Gillet, J.P., Varma, S. & Gottesman, M.M. The clinical relevance of cancer cell lines. *J Natl Cancer Inst* **105**, 452-458 (2013).
89. Neve, R.M., *et al.* A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515-527 (2006).
90. Tan, I.B., *et al.* Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology* **141**, 476-485, 485.e471-411 (2011).
91. Palanca-Wessels, M.C., *et al.* Extended lifespan of Barrett's esophagus epithelium transduced with the human telomerase catalytic subunit: a useful in vitro model. *Carcinogenesis* **24**, 1183-1190 (2003).
92. Zhang, X., *et al.* Malignant transformation of non-neoplastic Barrett's epithelial cells through well-defined genetic manipulations. *PLoS One* **5**(2010).
93. Ahuja, D., Sáenz-Robles, M.T. & Pipas, J.M. SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene* **24**, 7729-7745 (2005).
94. Underwood, T.J., *et al.* A comparison of primary oesophageal squamous epithelial cells with HET-1A in organotypic culture. *Biol Cell* **102**, 635-644 (2010).
95. Harada, H., *et al.* Telomerase induces immortalization of human esophageal keratinocytes without p16INK4a inactivation. *Mol Cancer Res* **1**, 729-738 (2003).
96. Jin, K., *et al.* Patient-derived human tumour tissue xenografts in immunodeficient mice: a systematic review. *Clin Transl Oncol* **12**, 473-480 (2010).

97. Bremner, C.G., Lynch, V.P. & Ellis, F.H. Barrett's esophagus: congenital or acquired? An experimental study of esophageal mucosal regeneration in the dog. *Surgery* **68**, 209-216 (1970).
98. Li, Y. & Martin, R.C. Reflux injury of esophageal mucosa: experimental studies in animal models of esophagitis, Barrett's esophagus and esophageal adenocarcinoma. *Dis Esophagus* **20**, 372-378 (2007).
99. Attwood, S.E.*, et al.* Duodenoesophageal reflux and the development of esophageal adenocarcinoma in rats. *Surgery* **111**, 503-510 (1992).
100. Attwood, S.E., Harrison, L.A., Preston, S.L. & Jankowski, J.A. Esophageal adenocarcinoma in "mice and men": back to basics! *Am J Gastroenterol* **103**, 2367-2372 (2008).
101. Wang, X.*, et al.* Residual embryonic cells as precursors of a Barrett's-like metaplasia. *Cell* **145**, 1023-1035 (2011).
102. Quante, M.*, et al.* Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. *Cancer Cell* **21**, 36-51 (2012).
103. Bird-Lieberman, E. & Fitzgerald, R. Early diagnosis of oesophageal cancer. *Br J Cancer* **101**, 1-6 (2009).
104. Graham, D.Y., Schwartz, J.T., Cain, G.D. & Gyorkey, F. Prospective evaluation of biopsy number in the diagnosis of esophageal and gastric carcinoma. *Gastroenterology* **82**, 228-231 (1982).
105. Siewert, J.R. & Stein, H.J. Classification of adenocarcinoma of the oesophagogastric junction. *Br J Surg* **85**, 1457-1459 (1998).
106. Low, D.E. Update on staging and surgical treatment options for esophageal cancer. *J Gastrointest Surg* **15**, 719-729 (2011).
107. Union Internationale Contre le Cancer. Oesophagus including Oesophagogastric Junction. in *TNM Classification of Malignant Tumours* (eds. Sobin, L.H., Gospodarowicz, M. & Wittekind, C.) 66-72 (Wiley-Blackwell, New York, 2009).
108. Pennathur, A., Gibson, M.K., Jobe, B.A. & Luketich, J.D. Oesophageal carcinoma. *Lancet* **381**, 400-412 (2013).
109. Wang, V.S., Hornick, J.L., Sepulveda, J.A., Mauer, R. & Poneros, J.M. Low prevalence of submucosal invasive carcinoma at esophagectomy for high-grade dysplasia or intramucosal adenocarcinoma in Barrett's esophagus: a 20-year experience. *Gastrointest Endosc* **69**, 777-783 (2009).
110. Shaheen, N.J.*, et al.* Durability of radiofrequency ablation in Barrett's esophagus with dysplasia. *Gastroenterology* **141**, 460-468 (2011).
111. Shaheen, N.J.*, et al.* Radiofrequency ablation in Barrett's esophagus with dysplasia. *N Engl J Med* **360**, 2277-2288 (2009).
112. Curvers, W.L., Bansal, A., Sharma, P. & Bergman, J.J. Endoscopic work-up of early Barrett's neoplasia. *Endoscopy* **40**, 1000-1007 (2008).
113. Mino-Kenudson, M.*, et al.* EMR for Barrett's esophagus-related superficial neoplasms offers better diagnostic reproducibility than mucosal biopsy. *Gastrointest Endosc* **66**, 660-666; quiz 767, 769 (2007).
114. Griffin, S.M., Burt, A.D. & Jennings, N.A. Lymph node metastasis in early esophageal adenocarcinoma. *Ann Surg* **254**, 731-736; discussion 736-737 (2011).
115. Leers, J.M.*, et al.* The prevalence of lymph node metastases in patients with T1 esophageal adenocarcinoma a retrospective review of esophagectomy specimens. *Ann Surg* **253**, 271-278 (2011).
116. Barbour, A.P.*, et al.* Risk stratification for early esophageal adenocarcinoma: analysis of lymphatic spread and prognostic factors. *Ann Surg Oncol* **17**, 2494-2502 (2010).
117. O'Neill, J.R.*, et al.* Defining a positive circumferential resection margin in oesophageal cancer and its implications for adjuvant treatment. *Br J Surg* **100**, 1055-1063 (2013).
118. Dresner, S.M., Lamb, P.J., Bennett, M.K., Hayes, N. & Griffin, S.M. The pattern of metastatic lymph node dissemination from adenocarcinoma of the esophagogastric junction. *Surgery* **129**, 103-109 (2001).
119. Cense, H.A.*, et al.* Lymphatic drainage routes of the gastric cardia visualized by lymphoscintigraphy. *J Nucl Med* **45**, 247-252 (2004).

120. Lerut, T. Esophageal surgery at the end of the millennium. *J Thorac Cardiovasc Surg* **116**, 1-20 (1998).
121. Hulscher, J.B., *et al.* Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the esophagus. *N Engl J Med* **347**, 1662-1669 (2002).
122. Gertler, R., *et al.* Long-term outcome of 2920 patients with cancers of the esophagus and esophagogastric junction: evaluation of the New Union Internationale Contre le Cancer/American Joint Cancer Committee staging system. *Ann Surg* **253**, 689-698 (2011).
123. Crabtree, T.D., *et al.* Evaluation of the reliability of clinical staging of T2 N0 esophageal cancer: a review of the Society of Thoracic Surgeons database. *Ann Thorac Surg* **96**, 382-390 (2013).
124. Allum, W.H., *et al.* Guidelines for the management of oesophageal and gastric cancer. *Gut* **60**, 1449-1472 (2011).
125. MRC Oesophageal Cancer Working Group. Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial. *Lancet* **359**, 1727-1733 (2002).
126. Markar, S.R., Karthikesalingam, A., Thrumurthy, S. & Low, D.E. Volume-outcome relationship in surgery for esophageal malignancy: systematic review and meta-analysis 2000-2011. *J Gastrointest Surg* **16**, 1055-1063 (2012).
127. Groene, O., *et al.* National Oesophago-Gastric Cancer Audit. (Royal College of Surgeons of England, 2012).
128. Kelsen, D.P., *et al.* Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer. *N Engl J Med* **339**, 1979-1984 (1998).
129. Cunningham, D., *et al.* Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* **355**, 11-20 (2006).
130. Alderson, D. http://www.cancer.gov/clinicaltrials/search/view?cdrid=69457&version=healthprofessional. Phase III Randomized Study of Neoadjuvant Cisplatin and Fluorouracil Versus Cisplatin, Epirubicin, and Fluorouracil in Patients With Resectable Adenocarcinoma of the Esophagus (2013).
131. Arnott, S.J., *et al.* Preoperative radiotherapy for esophageal carcinoma. *Cochrane Database Syst Rev*, CD001799 (2005).
132. Urschel, J.D. & Vasan, H. A meta-analysis of randomized controlled trials that compared neoadjuvant chemoradiation and surgery to surgery alone for resectable esophageal cancer. *Am J Surg* **185**, 538-543 (2003).
133. Bosset, J.F., *et al.* Chemoradiotherapy followed by surgery compared with surgery alone in squamous-cell cancer of the esophagus. *N Engl J Med* **337**, 161-167 (1997).
134. Sjoquist, K.M., *et al.* Survival after neoadjuvant chemotherapy or chemoradiotherapy for resectable oesophageal carcinoma: an updated meta-analysis. *Lancet Oncol* **12**, 681-692 (2011).
135. Morgan, M.A., *et al.* Stage-for-stage comparison of definitive chemoradiotherapy, surgery alone and neoadjuvant chemotherapy for oesophageal carcinoma. *Br J Surg* **96**, 1300-1307 (2009).
136. Mooney, M. Neoadjuvant and adjuvant chemotherapy for esophageal adenocarcinoma. *Journal of Surgical Oncology* **92**, 230-238 (2005).
137. Reynolds, J.V., *et al.* Long-term outcomes following neoadjuvant chemoradiotherapy for esophageal cancer. *Ann Surg* **245**, 707-716 (2007).
138. Courrech Staal, E.F., *et al.* Systematic review of the benefits and risks of neoadjuvant chemoradiation for oesophageal cancer. *Br J Surg* **97**, 1482-1496 (2010).
139. Power, D.G. & Reynolds, J.V. Localized adenocarcinoma of the esophagogastric junction--is there a standard of care? *Cancer Treat Rev* **36**, 400-409 (2010).
140. Kelsen, D.P., *et al.* Long-term results of RTOG trial 8911 (USA Intergroup 113): a random assignment trial comparison of chemotherapy followed by surgery compared with surgery alone for esophageal cancer. *J Clin Oncol* **25**, 3719-3725 (2007).
141. Courrech Staal, E.F., *et al.* Systematic review of the benefits and risks of neoadjuvant chemoradiation for oesophageal cancer. *Br J Surg* **97**, 1482-1496 (2010).
142. Goldfarb, Y., *et al.* Improving postoperative immune status and resistance to cancer metastasis: a combined perioperative approach of immunostimulation and prevention of excessive surgical stress responses. *Ann Surg* **253**, 798-810 (2011).

143. Grotenhuis, B.A., *et al.* Lymphatic micrometastases in patients with early esophageal adenocarcinoma. *J Surg Oncol* **102**, 863-867 (2010).
144. Vashist, Y.K., *et al.* Disseminated tumor cells in bone marrow and the natural course of resected esophageal cancer. *Ann Surg* **255**, 1105-1112 (2012).
145. Schmidt-Kittler, O., *et al.* From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc Natl Acad Sci U S A* **100**, 7737-7742 (2003).
146. Marches, R., Scheuermann, R. & Uhr, J. Cancer dormancy: from mice to man. *Cell Cycle* **5**, 1772-1778 (2006).
147. Matsuzawa, A., Takeda, Y., Narita, M. & Ozawa, H. Survival of leukemic cells in a dormant state following cyclophosphamide-induced cure of strongly immunogenic mouse leukemia (DL811). *Int J Cancer* **49**, 303-309 (1991).
148. Hüsemann, Y., *et al.* Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58-68 (2008).
149. Bajaj, N.S., *et al.* Donor transmission of malignant melanoma in a lung transplant recipient 32 years after curative resection. *Transpl Int* **23**, e26-31 (2010).
150. Patani, N., *et al.* Direct evidence for concurrent morphological and genetic heterogeneity in an invasive ductal carcinoma of triple-negative phenotype. *J Clin Pathol* **64**, 822-828 (2011).
151. Goldfarb, Y., *et al.* Improving postoperative immune status and resistance to cancer metastasis: a combined perioperative approach of immunostimulation and prevention of excessive surgical stress responses. *Ann Surg* **253**, 798-810 (2011).
152. Weinstein, I.B. Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science* **297**, 63-64 (2002).
153. Druker, B.J., *et al.* Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* **2**, 561-566 (1996).
154. Shtivelman, E., Lifshitz, B., Gale, R.P. & Canaani, E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. *Nature* **315**, 550-554 (1985).
155. Hughes, T.P., *et al.* Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N Engl J Med* **349**, 1423-1432 (2003).
156. Kris, M.G., *et al.* Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* **290**, 2149-2158 (2003).
157. Sordella, R., Bell, D.W., Haber, D.A. & Settleman, J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* **305**, 1163-1167 (2004).
158. Grünwald, V. & Hidalgo, M. Development of the epidermal growth factor receptor inhibitor Tarceva (OSI-774). *Adv Exp Med Biol* **532**, 235-246 (2003).
159. Davies, H., *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).
160. Flaherty, K.T., *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* **363**, 809-819 (2010).
161. Chapman, P.B., *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* **364**, 2507-2516 (2011).
162. Piccart-Gebhart, M.J., *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* **353**, 1659-1672 (2005).
163. Gravalos, C. & Jimeno, A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Ann Oncol* **19**, 1523-1529 (2008).
164. Reichelt, U., *et al.* Frequent homogeneous HER-2 amplification in primary and metastatic adenocarcinoma of the esophagus. *Mod Pathol* **20**, 120-129 (2007).
165. Bang, Y.J., *et al.* Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* **376**, 687-697 (2010).

166. Safran, H., *et al.* Phase I/II study of trastuzumab, paclitaxel, cisplatin and radiation for locally advanced, HER2 overexpressing, esophageal adenocarcinoma. *Int J Radiat Oncol Biol Phys* **67**, 405-409 (2007).
167. Iqbal, S., *et al.* Southwest Oncology Group study S0413: a phase II trial of lapatinib (GW572016) as first-line therapy in patients with advanced or metastatic gastric cancer. *Ann Oncol* **22**, 2610-2615 (2011).
168. Dragovich, T., *et al.* Phase II trial of erlotinib in gastroesophageal junction and gastric adenocarcinomas: SWOG 0127. *J Clin Oncol* **24**, 4922-4927 (2006).
169. Greenwalt, D.M., *et al.* Gene expression profiling of esophageal cancer: Comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma. *International Journal of Cancer* **120**, 1914-1921 (2007).
170. Nancarrow, D.J., *et al.* Whole genome expression array profiling highlights differences in mucosal defense genes in Barrett's esophagus and esophageal adenocarcinoma. *PLoS One* **6**, e22513 (2011).
171. Hao, Y., *et al.* Gene Expression Profiling reveals Stromal Genes Expressed in Common Between Barrett's Esophagus and Adenocarcinoma. *Gastroenterology* **131**, 925-933 (2006).
172. Kaelin, W.G. The concept of synthetic lethality in the context of anticancer therapy. *Nature Reviews Cancer* **5**, 689-698 (2005).
173. Lane, D.P. & Crawford, L.V. T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278**, 261-263 (1979).
174. Isobe, M., Emanuel, B.S., Givol, D., Oren, M. & Croce, C.M. Localization of gene for human p53 tumour antigen to band 17p13. *Nature* **320**, 84-85 (1986).
175. Laptenko, O. & Prives, C. Transcriptional regulation by p53: one protein, many possibilities. *Cell Death Differ* **13**, 951-961 (2006).
176. Parada, L.F., Land, H., Weinberg, R.A., Wolf, D. & Rotter, V. Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation. *Nature* **312**, 649-651 (1984).
177. Jenkins, J.R., Rudge, K., Chumakov, P. & Currie, G.A. The cellular oncogene p53 can be activated by mutagenesis. *Nature* **317**, 816-818 (1985).
178. Finlay, C.A., Hinds, P.W. & Levine, A.J. The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**, 1083-1093 (1989).
179. Wei, C.L., *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207-219 (2006).
180. Lane, D. Cancer. p53, guardian of the genome. *Nature* **358**, 15-16 (1992).
181. Vogelstein, B., Lane, D. & Levine, A. Surfing the p53 network. *Nature* **408**, 307-310 (2000).
182. Raycroft, L., Wu, H.Y. & Lozano, G. Transcriptional activation by wild-type but not transforming mutants of the p53 anti-oncogene. *Science* **249**, 1049-1051 (1990).
183. Campbell, H.G., *et al.* Activation of p53 following ionizing radiation, but not other stressors, is dependent on the proline-rich domain (PRD). *Oncogene* **32**, 827-836 (2013).
184. MacLaine, N., *et al.* A central role for CK1 in catalyzing phosphorylation of the p53 transactivation domain at serine 20 after HHV-6B viral infection. *J Biol Chem* **283**, 28563-28573 (2008).
185. el-Deiry, W.S., Kern, S.E., Pietenpol, J.A., Kinzler, K.W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nat Genet* **1**, 45-49 (1992).
186. Olivier, M., Hollstein, M. & Hainaut, P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol* **2**, a001008 (2010).
187. Clore, G.M., *et al.* High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* **265**, 386-391 (1994).
188. Lomax, M., Barnes, D., Hupp, T., Picksley, S. & Camplejohn, R. Characterization of p53 oligomerization domain mutations isolated from Li-Fraumeni and Li-Fraumeni like family members. *Oncogene* **17**, 643-649 (1998).
189. Hupp, T.R., Meek, D.W., Midgley, C.A. & Lane, D.P. Regulation of the specific DNA binding function of p53. *Cell* **71**, 875-886 (1992).
190. Hupp, T., Sparks, A. & Lane, D. Small peptides activate the latent sequence-specific DNA binding function of p53. *Cell* **83**, 237-245 (1995).

191. Momand, J., Zambetti, G.P., Olson, D.C., George, D. & Levine, A.J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237-1245 (1992).
192. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387**, 296-299 (1997).
193. Nenutil, R.*, et al.* Discriminating functional and non-functional p53 in human tumours by p53 and MDM2 immunohistochemistry. *J Pathol* **207**, 251-259 (2005).
194. Kandoth, C.*, et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73 (2013).
195. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615 (2011).
196. Stransky, N.*, et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157-1160 (2011).
197. Agrawal, N.*, et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**, 1154-1157 (2011).
198. Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J. & Howley, P.M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129-1136 (1990).
199. Leedham, S.J.*, et al.* Individual crypt genetic heterogeneity and the origin of metaplastic glandular epithelium in human Barrett's oesophagus. *Gut* **57**, 1041-1048 (2008).
200. Gaiddon, C., Lokshin, M., Ahn, J., Zhang, T. & Prives, C. A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain. *Molecular and Cellular Biology* **21**, 1874-1887 (2001).
201. Kravchenko, J.*, et al.* Small-molecule RETRA suppresses mutant p53-bearing cancer cells through a p73-dependent salvage pathway. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6302-6307 (2008).
202. Milner, J. & Medcalf, E.A. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. *Cell* **65**, 765-774 (1991).
203. Bossi, G.*, et al.* Mutant p53 gain of function: reduction of tumor malignancy of human cancer cell lines through abrogation of mutant p53 expression. *Oncogene* **25**, 304-309 (2006).
204. Adorno, M.*, et al.* A Mutant-p53/Smad complex opposes p63 to empower TGF beta-induced metastasis. *Cell* **137**, 87-98 (2009).
205. Bykov, V.*, et al.* Restoration of the tumor suppressor function to mutant p53 by a low-molecular-weight compound. *Nature Medicine* **8**, 282-283 (2002).
206. Olivier, M.*, et al.* The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat* **19**, 607-614 (2002).
207. Wang, Z. & Sun, Y. Targeting p53 for Novel Anticancer Therapy. *Translational Oncology* **3**, 1-12 (2010).
208. Madani, K., Zhao, R., Lim, H.J. & Casson, A.G. Prognostic value of p53 mutations in oesophageal adenocarcinoma: final results of a 15-year prospective study. *Eur J Cardiothorac Surg* **37**, 1427-1432 (2010).
209. Russo, A.*, et al.* p53 mutations in L3-loop zinc-binding domain, DNA-ploidy, and S phase fraction are independent prognostic indicators in colorectal cancer: a prospective study with a five-year follow-up. *Cancer Epidemiology Biomarkers and Prevention* **11**, 1322-1331 (2002).
210. Kato, S.*, et al.* Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 8424-8429 (2003).
211. Sanger Institute. COSMIC, The Catalogue of Somatic Mutations in Cancer Database. (2010).
212. Hamelin, R.*, et al.* TP53 gene mutations and p53 protein immunoreactivity in malignant and premalignant Barrett's esophagus. *Gastroenterology* **107**, 1012-1018 (1994).
213. Ireland, A.*, et al.* Clinical significance of p53 mutations in adenocarcinoma of the esophagus and cardia. *Ann Surg* **231**, 179-187 (2000).

214. Halm, U., *et al.* Apoptosis and cell proliferation in the metaplasia-dysplasia-carcinoma-sequence of Barrett's esophagus. *Hepatogastroenterology* **47**, 962-966 (2000).
215. Bian, Y.S., Osterheld, M.C., Bosman, F.T., Benhattar, J. & Fontolliet, C. p53 gene mutation and protein accumulation during neoplastic progression in Barrett's esophagus. *Mod Pathol* **14**, 397-403 (2001).
216. Chung, S., Kao, J., Hyjek, E. & Chen, Y. p53 in esophageal adenocarcinoma: a critical reassessment of mutation frequency and identification of 72Arg as the dominant allele. *International Journal of Oncology* **31**, 1351-1355 (2007).
217. Doak, S., *et al.* Characterisation of p53 status at the gene, chromosomal and protein levels in oesophageal adenocarcinoma. *British Journal of Cancer* **89**, 1729-1735 (2003).
218. Dulak, A.M., *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* **45**, 478-486 (2013).
219. Ashworth, A. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol* **26**, 3785-3790 (2008).
220. Sur, S., *et al.* A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proceedings on the National Academy of Science of the United States of America* **106**, 3964-3969 (2009).
221. Vaske, C.J., *et al.* Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics* **26**, i237-245 (2010).
222. Wierstra, I. & Alves, J. FOXM1, a typical proliferation-associated transcription factor. *Biol Chem* **388**, 1257-1274 (2007).
223. Hartwell, L.H. & Weinert, T.A. Checkpoints: controls that ensure the order of cell cycle events. *Science* **246**, 629-634 (1989).
224. Bernstein, C., Bernstein, H., Payne, C. & Garewal, H. DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutat Res* **511**, 145-178 (2002).
225. De Witt Hamer, P.C., Mir, S.E., Noske, D., Van Noorden, C.J. & Würdinger, T. WEE1 kinase targeting combined with DNA-damaging cancer therapy catalyzes mitotic catastrophe. *Clin Cancer Res* **17**, 4200-4207 (2011).
226. Mir, S.E., *et al.* In silico analysis of kinase expression identifies WEE1 as a gatekeeper against mitotic catastrophe in glioblastoma. *Cancer Cell* **18**, 244-257 (2010).
227. Kuntz, K. & O'Connell, M.J. The G(2) DNA damage checkpoint: could this ancient regulator be the Achilles heel of cancer? *Cancer Biol Ther* **8**, 1433-1439 (2009).
228. Llamazares, S., *et al. polo* encodes a protein kinase homolog required for mitosis in *Drosophila*. *Genes Dev* **5**, 2153-2165 (1991).
229. Sunkel, C.E. & Glover, D.M. *polo*, a mitotic mutant of *Drosophila* displaying abnormal spindle poles. *J Cell Sci* **89 (Pt 1)**, 25-38 (1988).
230. Hamanaka, R., *et al.* Cloning and characterization of human and murine homologues of the *Drosophila polo* serine-threonine kinase. *Cell Growth Differ* **5**, 249-257 (1994).
231. Lowery, D.M., Lim, D. & Yaffe, M.B. Structure and function of Polo-like kinases. *Oncogene* **24**, 248-259 (2005).
232. Elia, A.E., *et al.* The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* **115**, 83-95 (2003).
233. de Cárcer, G., *et al.* Plk5, a polo box domain-only protein with specific roles in neuron differentiation and glioblastoma suppression. *Mol Cell Biol* **31**, 1225-1239 (2011).
234. Nigg, E.A. Mitotic kinases as regulators of cell division and its checkpoints. *Nat Rev Mol Cell Biol* **2**, 21-32 (2001).
235. Bassermann, F., *et al.* The Cdc14B-Cdh1-Plk1 axis controls the G2 DNA-damage-response checkpoint. *Cell* **134**, 256-267 (2008).
236. Macurek, L., *et al.* Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* **455**, 119-123 (2008).
237. Macurek, L., Lindqvist, A. & Medema, R. Aurora-A and hBora join the game of Polo. *Cancer Res* **69**, 4555-4558 (2009).

238. Seki, A., Coppinger, J., Jang, C., Yates, J. & Fang, G. Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. *Science* **320**, 1655-1658 (2008).
239. Lane, H.A. & Nigg, E.A. Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes. *J Cell Biol* **135**, 1701-1713 (1996).
240. van Vugt, M.A., Brás, A. & Medema, R.H. Polo-like kinase-1 controls recovery from a G2 DNA damage-induced arrest in mammalian cells. *Mol Cell* **15**, 799-811 (2004).
241. Kang, Y.H., *et al.* Self-regulated Plk1 recruitment to kinetochores by the Plk1-PBIP1 interaction is critical for proper chromosome segregation. *Mol Cell* **24**, 409-422 (2006).
242. Burkard, M.E., *et al.* Chemical genetics reveals the requirement for Polo-like kinase 1 activity in positioning RhoA and triggering cytokinesis in human cells. *Proc Natl Acad Sci U S A* **104**, 4383-4388 (2007).
243. Neef, R., *et al.* Phosphorylation of mitotic kinesin-like protein 2 by polo-like kinase 1 is required for cytokinesis. *J Cell Biol* **162**, 863-875 (2003).
244. Seong, Y.S., *et al.* A spindle checkpoint arrest and a cytokinesis failure by the dominant-negative polo-box domain of Plk1 in U-2 OS cells. *J Biol Chem* **277**, 32282-32293 (2002).
245. van Vugt, M.A. & Medema, R.H. Getting in and out of mitosis with Polo-like kinase-1. *Oncogene* **24**, 2844-2859 (2005).
246. Leonard, M.K., Hill, N.T., Bubulya, P.A. & Kadakia, M.P. The PTEN-Akt pathway impacts the integrity and composition of mitotic centrosomes. *Cell Cycle* **12**, 1406-1415 (2013).
247. Uchiumi, T., Longo, D.L. & Ferris, D.K. Cell cycle regulation of the human polo-like kinase (PLK) promoter. *J Biol Chem* **272**, 9166-9174 (1997).
248. Krupczak-Hollis, K., *et al.* The mouse Forkhead Box m1 transcription factor is essential for hepatoblast mitosis and development of intrahepatic bile ducts and vessels during liver morphogenesis. *Dev Biol* **276**, 74-88 (2004).
249. Fu, Z., *et al.* Plk1-dependent phosphorylation of FoxM1 regulates a transcriptional programme required for mitotic progression. *Nat Cell Biol* **10**, 1076-1082 (2008).
250. Darieva, Z., *et al.* Polo kinase controls cell-cycle-dependent transcription by targeting a coactivator protein. *Nature* **444**, 494-498 (2006).
251. Ferris, D.K., Maloid, S.C. & Li, C.C. Ubiquitination and proteasome mediated degradation of polo-like kinase. *Biochem Biophys Res Commun* **252**, 340-344 (1998).
252. Jang, Y.J., Ma, S., Terada, Y. & Erikson, R.L. Phosphorylation of threonine 210 and the role of serine 137 in the regulation of mammalian polo-like kinase. *J Biol Chem* **277**, 44115-44120 (2002).
253. Yamashiro, S., *et al.* Myosin phosphatase-targeting subunit 1 regulates mitosis by antagonizing polo-like kinase 1. *Dev Cell* **14**, 787-797 (2008).
254. Syljuåsen, R., Jensen, S., Bartek, J. & Lukas, J. Adaptation to the ionizing radiation-induced G2 checkpoint occurs in human cells and depends on checkpoint kinase 1 and Polo-like kinase 1 kinases. *Cancer Res* **66**, 10253-10257 (2006).
255. Kops GJ, W.B., Cleveland DW. On the road to cancer: aneuploidy and the mitotic chechpoint. *Nature Reviews Cancer* **5**, 773-785 (2005).
256. Chen, Z., *et al.* Selective Chk1 inhibitors differentially sensitize p53-deficient cancer cells to cancer therapeutics. *Int J Cancer* **119**, 2784-2794 (2006).
257. Hirose, Y., Berger, M.S. & Pieper, R.O. p53 effects both the duration of G2/M arrest and the fate of temozolomide-treated human glioblastoma cells. *Cancer Res* **61**, 1957-1963 (2001).
258. Boulaire, J., Fotedar, A. & Fotedar, R. The functions of the cdk-cyclin kinase inhibitor p21WAF1. *Pathol Biol (Paris)* **48**, 190-202 (2000).
259. Hirai, H., *et al.* MK-1775, a small molecule Wee1 inhibitor, enhances anti-tumor efficacy of various DNA-damaging agents, including 5-fluorouracil. *Cancer Biol Ther* **9**(2010).
260. Cogswell, J.P., Brown, C.E., Bisi, J.E. & Neill, S.D. Dominant-negative polo-like kinase 1 induces mitotic catastrophe independent of cdc25C function. *Cell Growth Differ* **11**, 615-623 (2000).

261. Seki, A., *et al.* Plk1- and beta-TrCP-dependent degradation of Bora controls mitotic progression. *J Cell Biol* **181**, 65-78 (2008).
262. Holtrich, U., *et al.* Induction and down-regulation of PLK, a human serine/threonine kinase expressed in proliferating cells and tumors. *Proc Natl Acad Sci U S A* **91**, 1736-1740 (1994).
263. Yuan, J., *et al.* Polo-like kinase, a novel marker for cellular proliferation. *Am J Pathol* **150**, 1165-1172 (1997).
264. King, S.I., *et al.* Immunohistochemical detection of Polo-like kinase-1 (PLK1) in primary breast cancer is associated with TP53 mutation and poor clinical outcom. *Breast Cancer Res* **14**, R40 (2012).
265. Wolf, G., *et al.* Prognostic significance of polo-like kinase (PLK) expression in non-small cell lung cancer. *Oncogene* **14**, 543-549 (1997).
266. Knecht, R., *et al.* Prognostic significance of polo-like kinase (PLK) expression in squamous cell carcinomas of the head and neck. *Cancer Res* **59**, 2794-2797 (1999).
267. Smith, M.R., *et al.* Malignant transformation of mammalian cells initiated by constitutive expression of the polo-like kinase. *Biochem Biophys Res Commun* **234**, 397-405 (1997).
268. Tokumitsu, Y., *et al.* Prognostic significance of polo-like kinase expression in esophageal carcinoma. *Int J Oncol* **15**, 687-692 (1999).
269. Feng YB, L.D., Shi ZZ, Wang XC, Shen XM, Zhang Y, Du XL, Luo ML, Xu X, Han YL, Cai Y, Zhang ZQ, Zhan QM, Wang MR. Overexpression of PLK1 is associated with poor survival by inhibiting apoptosis via enhancement of survivin level in esophageal squamous cell carcinoma. *International Journal of Cancer* **124**, 578-588 (2009).
270. Ito, T., *et al.* Polo-like kinase 1 regulates cell proliferation and is targeted by miR-593* in esophageal cancer. *Int J Cancer* **129**, 2134-2146 (2011).
271. Sato, F., *et al.* Polo-like kinase and survivin are esophageal tumor-specific promoters. *Biochem Biophys Res Commun* **342**, 465-471 (2006).
272. Zhao, C., Gong, L., Li, W. & Chen, L. Overexpression of Plk1 promotes malignant progress in human esophageal squamous cell carcinoma. *J Cancer Res Clin Oncol* **136**, 9-16 (2010).
273. Bandla, S., *et al.* Comparative genomics of esophageal adenocarcinoma and squamous cell carcinoma. *Ann Thorac Surg* **93**, 1101-1106 (2012).
274. Bonde, P., *et al.* Cytogenetic characterization and gene expression profiling in the rat reflux-induced esophageal tumor model. *J Thorac Cardiovasc Surg* **133**, 763-769 (2007).
275. Peters, C.J., *et al.* A 4-gene signature predicts survival of patients with resected adenocarcinoma of the esophagus, junction, and gastric cardia. *Gastroenterology* **139**, 1995-2004.e1915 (2010).
276. Dibb, M., *et al.* The FOXM1-PLK1 axis is commonly upregulated in oesophageal adenocarcinoma. *Br J Cancer* **107**, 1766-1775 (2012).
277. Kim, S.M., *et al.* Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. *PLoS One* **5**, e15074 (2010).
278. Spänkuch-Schmitt, B., *et al.* Downregulation of human polo-like kinase activity by antisense oligonucleotides induces growth inhibition in cancer cells. *Oncogene* **21**, 3162-3171 (2002).
279. Spänkuch-Schmitt, B., Bereiter-Hahn, J., Kaufmann, M. & Strebhardt, K. Effect of RNA silencing of polo-like kinase-1 (PLK1) on apoptosis and spindle formation in human cancer cells. *J Natl Cancer Inst* **94**, 1863-1877 (2002).
280. Elez, R., *et al.* Tumor regression by combination antisense therapy against Plk1 and Bcl-2. *Oncogene* **22**, 69-80 (2003).
281. Guan, R., *et al.* Small interfering RNA-mediated Polo-like kinase 1 depletion preferentially reduces the survival of p53-defective, oncogenic transformed cells and inhibits tumor growth in animals. *Cancer Res* **65**, 2698-2704 (2005).
282. Liu, X. & Erikson, R.L. Activation of Cdc2/cyclin B and inhibition of centrosome amplification in cells depleted of Plk1 by siRNA. *Proc Natl Acad Sci U S A* **99**, 8672-8676 (2002).
283. Liu, X. & Erikson, R.L. Polo-like kinase (Plk)1 depletion induces apoptosis in cancer cells. *Proc Natl Acad Sci U S A* **100**, 5789-5794 (2003).

284. van Vugt, M., Brás, A. & Medema, R. Polo-like kinase-1 controls recovery from a G2 DNA damage-induced arrest in mammalian cells. *Molecular Cell* **15**, 799-811 (2004).
285. van Vugt, M.A., *et al.* Polo-like kinase-1 is required for bipolar spindle formation but is dispensable for anaphase promoting complex/Cdc20 activation and initiation of cytokinesis. *J Biol Chem* **279**, 36841-36854 (2004).
286. Sumara, I., *et al.* Roles of polo-like kinase 1 in the assembly of functional mitotic spindles. *Curr Biol* **14**, 1712-1722 (2004).
287. Johnson, E.F., Stewart, K.D., Woods, K.W., Giranda, V.L. & Luo, Y. Pharmacological and functional comparison of the polo-like kinase family: insight into inhibitor and substrate specificity. *Biochemistry* **46**, 9551-9563 (2007).
288. Kothe, M., *et al.* Selectivity-determining residues in Plk1. *Chem Biol Drug Des* **70**, 540-546 (2007).
289. Kothe, M., *et al.* Structure of the catalytic domain of human polo-like kinase 1. *Biochemistry* **46**, 5960-5971 (2007).
290. Steegmaier, M., *et al.* BI 2536, a potent and selective inhibitor of polo-like kinase 1, inhibits tumor growth in vivo. *Curr Biol* **17**, 316-322 (2007).
291. Donohue, P.J., Alberts, G.F., Guo, Y. & Winkles, J.A. Identification by targeted differential display of an immediate early gene encoding a putative serine/threonine kinase. *J Biol Chem* **270**, 10351-10357 (1995).
292. Simmons, D.L., Neel, B.G., Stevens, R., Evett, G. & Erikson, R.L. Identification of an early-growth-response gene encoding a novel putative protein kinase. *Mol Cell Biol* **12**, 4164-4169 (1992).
293. Smith, P., Syed, N. & Crook, T. Epigenetic inactivation implies a tumor suppressor function in hematologic malignancies for Polo-like kinase 2 but not Polo-like kinase 3. *Cell Cycle* **5**, 1262-1264 (2006).
294. Xie, S., *et al.* Plk3 functionally links DNA damage to cell cycle arrest and apoptosis at least in part via the p53 pathway. *J Biol Chem* **276**, 43305-43312 (2001).
295. Gumireddy, K., *et al.* ON01910, a non-ATP-competitive small molecule inhibitor of Plk1, is a potent anticancer agent. *Cancer Cell* **7**, 275-286 (2005).
296. McInnes, C., *et al.* Inhibitors of Polo-like kinase reveal roles in spindle-pole maintenance. *Nat Chem Biol* **2**, 608-617 (2006).
297. Peters, U., Cherian, J., Kim, J.H., Kwok, B.H. & Kapoor, T.M. Probing cell-division phenotype space and Polo-like kinase function using small molecules. *Nat Chem Biol* **2**, 618-626 (2006).
298. Reindl, W., Yuan, J., Krämer, A., Strebhardt, K. & Berg, T. Inhibition of polo-like kinase 1 by blocking polo-box domain-dependent protein-protein interactions. *Chem Biol* **15**, 459-466 (2008).
299. Watanabe, N., *et al.* Deficiency in chromosome congression by the inhibition of Plk1 polo box domain-dependent recognition. *J Biol Chem* **284**, 2344-2353 (2009).
300. Reindl, W., Yuan, J., Krämer, A., Strebhardt, K. & Berg, T. A pan-specific inhibitor of the polo-box domains of polo-like kinases arrests cancer cells in mitosis. *Chembiochem* **10**, 1145-1148 (2009).
301. Hanisch, A., Wehner, A., Nigg, E.A. & Silljé, H.H. Different Plk1 functions show distinct dependencies on Polo-Box domain-mediated targeting. *Mol Biol Cell* **17**, 448-459 (2006).
302. Strebhardt, K. Multifaceted polo-like kinases: drug targets and antitargets for cancer therapy. *Nat Rev Drug Discov* **9**, 643-660 (2010).
303. Rudolph, D., *et al.* BI 6727, a Polo-like kinase inhibitor with improved pharmacokinetic profile and broad antitumor activity. *Clin Cancer Res* **15**, 3094-3102 (2009).
304. Sato, Y., *et al.* Imidazopyridine derivatives as potent and selective Polo-like kinase (PLK) inhibitors. *Bioorg Med Chem Lett* **19**, 4673-4678 (2009).
305. Emmitte, K.A., *et al.* Discovery of thiophene inhibitors of polo-like kinase. *Bioorg Med Chem Lett* **19**, 1018-1021 (2009).
306. Olmos, D., *et al.* Phase I study of GSK461364, a specific and competitive Polo-like kinase 1 inhibitor, in patients with advanced solid malignancies. *Clin Cancer Res* **17**, 3420-3430 (2011).

307. Tanaka, H., *et al.* HMN-176, an active metabolite of the synthetic antitumor agent HMN-214, restores chemosensitivity to multidrug-resistant cells by targeting the transcription factor NF-Y. *Cancer Res* **63**, 6942-6947 (2003).
308. Garland, L.L., Taylor, C., Pilkington, D.L., Cohen, J.L. & Von Hoff, D.D. A phase I pharmacokinetic study of HMN-214, a novel oral stilbene derivative with polo-like kinase-1-interacting properties, in patients with advanced solid tumors. *Clin Cancer Res* **12**, 5182-5189 (2006).
309. Uckun, F.M., *et al.* Anti-breast cancer activity of LFM-A13, a potent inhibitor of Polo-like kinase (PLK). *Bioorg Med Chem* **15**, 800-814 (2007).
310. Valsasina, B., *et al.* NMS-P937, an orally available, specific small-molecule polo-like kinase 1 inhibitor with antitumor activity in solid and hematologic malignancies. *Mol Cancer Ther* **11**, 1006-1016 (2012).
311. Keppner, S., *et al.* Biological impact of freezing Plk1 in its inactive conformation in cancer cells. *Cell Cycle* **9**, 761-773 (2010).
312. Santamaria, A., *et al.* Use of the novel Plk1 inhibitor ZK-thiazolidinone to elucidate functions of Plk1 in early and late stages of mitosis. *Mol Biol Cell* **18**, 4024-4036 (2007).
313. Mross, K., *et al.* Phase I dose escalation and pharmacokinetic study of BI 2536, a novel Polo-like kinase 1 inhibitor, in patients with advanced solid tumors. *J Clin Oncol* **26**, 5511-5517 (2008).
314. Hofheinz, R.D., *et al.* An open-label, phase I study of the polo-like kinase-1 inhibitor, BI 2536, in patients with advanced solid tumors. *Clin Cancer Res* **16**, 4666-4674 (2010).
315. Frost, A., *et al.* Phase I study of the Plk1 inhibitor BI 2536 administered intravenously on three consecutive days in advanced solid tumours. *Curr Oncol* **19**, e28-35 (2012).
316. Ellis, P.M., *et al.* A phase I open-label dose-escalation study of intravenous BI 2536 together with pemetrexed in previously treated patients with non-small-cell lung cancer. *Clin Lung Cancer* **14**, 19-27 (2013).
317. Sebastian, M., *et al.* The efficacy and safety of BI 2536, a novel Plk-1 inhibitor, in patients with stage IIIB/IV non-small cell lung cancer who had relapsed after, or failed, chemotherapy: results from an open-label, randomized phase II clinical trial. *J Thorac Oncol* **5**, 1060-1067 (2010).
318. Schöffski, P., *et al.* Multicentric parallel phase II trial of the polo-like kinase 1 inhibitor BI 2536 in patients with advanced head and neck cancer, breast cancer, ovarian cancer, soft tissue sarcoma and melanoma. The first protocol of the European Organization for Research and Treatment of Cancer (EORTC) Network Of Core Institutes (NOCI). *Eur J Cancer* **46**, 2206-2215 (2010).
319. Mross, K., *et al.* A randomised phase II trial of the Polo-like kinase inhibitor BI 2536 in chemo-naïve patients with unresectable exocrine adenocarcinoma of the pancreas - a study within the Central European Society Anticancer Drug Research (CESAR) collaborative network. *Br J Cancer* **107**, 280-286 (2012).
320. Pandha, H., *et al.* An open label phase II trial of BI 2536, a novel PLK1 inhibitor, in patients with metastatic hormone refractory prostate cancer (HRPC). in *American Society of Clinical Oncology*, Vol. 26 14547 (Journal of Clinical Oncology, 2008).
321. Jimeno, A., *et al.* Phase I study of ON 01910.Na, a novel modulator of the Polo-like kinase 1 pathway, in adult patients with solid tumors. *J Clin Oncol* **26**, 5504-5510 (2008).
322. Vainshtein, J., *et al.* Phase I study of ON 01910.Na, a novel polo-like kinase 1 pathway modulator, administered as a weekly 24-hour continuous infusion in patients with advanced cancer. in *American Society of Clinical Oncology*, Vol. 26 2515 (2008).
323. Ohnuma, T., *et al.* Phase I study of ON 01910.Na by 3-day continuous infusion (CI) in patients (pts) with advanced cancer. in *American Society of Clinical Oncology*, Vol. 24 13137 (2006).
324. Ma, W.W., *et al.* Phase I study of Rigosertib, an inhibitor of the phosphatidylinositol 3-kinase and Polo-like kinase 1 pathways, combined with gemcitabine in patients with solid tumors and pancreatic cancer. *Clin Cancer Res* **18**, 2048-2055 (2012).
325. Olnes, M.J., *et al.* Directed therapy for patients with myelodysplastic syndromes (MDS) by suppression of cyclin D1 with ON 01910.Na. *Leuk Res* **36**, 982-989 (2012).
326. Raza, A., Mukherjee, S., Eisenberger, A., Mears, J. & Wilhelm, F. Phase II study of orally administered rigosertib (ON 01910.Na) in transfusion dependent lower-risk

myelodysplastic syndrome (MDS) patients. in *American Society for Clinical Oncology*, Vol. 31 7031 (Journal of Clinical Oncology, 2013).
327. Schöffski, P. Polo-like kinase (PLK) inhibitors in preclinical and early clinical development in oncology. *Oncologist* **14**, 559-570 (2009).
328. Schöffski, P., *et al.* A phase I, dose-escalation study of the novel Polo-like kinase inhibitor volasertib (BI 6727) in patients with advanced solid tumours. *Eur J Cancer* **48**, 179-186 (2012).
329. Lu, L.Y., *et al.* Polo-like kinase 1 is essential for early embryonic development and tumor suppression. *Mol Cell Biol* **28**, 6870-6876 (2008).
330. Lu, B., *et al.* The Plk1 inhibitor BI 2536 temporarily arrests primary cardiac fibroblasts in mitosis and generates aneuploidy in vitro. *PLoS One* **5**, e12963 (2010).
331. Liu, X., Lei, M. & Erikson, R.L. Normal cells, but not cancer cells, survive severe Plk1 depletion. *Mol Cell Biol* **26**, 2093-2108 (2006).
332. Donaldson, M.M., Tavares, A.A., Ohkura, H., Deak, P. & Glover, D.M. Metaphase arrest with centromere separation in polo mutants of Drosophila. *J Cell Biol* **153**, 663-676 (2001).
333. Liu, X., *et al.* CCT chaperonin complex is required for the biogenesis of functional Plk1. *Mol Cell Biol* **25**, 4993-5010 (2005).
334. Raab, M., *et al.* Toxicity modelling of Plk1-targeted therapies in genetically engineered mice and cultured primary mammalian cells. *Nat Commun* **2**, 395 (2011).
335. Herrmann, S., Amorim, I. & Sunkel, C.E. The *polo* kinase is required at multiple stages during spermatogenesis in *Drosophila melanogaster*. *Chromosoma* **107**, 440-451 (1998).
336. Karnoub, A.E. & Weinberg, R.A. Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol* **9**, 517-531 (2008).
337. Luo, J., *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848 (2009).
338. Sanhaji, M., *et al.* p53 is not directly relevant to the response of Polo-like kinase 1 inhibitors. *Cell Cycle* **11**, 543-553 (2012).
339. Kho, P.S., *et al.* p53-regulated transcriptional program associated with genotoxic stress-induced apoptosis. *J Biol Chem* **279**, 21183-21192 (2004).
340. McKenzie, L., *et al.* p53-dependent repression of polo-like kinase-1 (PLK1). *Cell Cycle* **9**, 4200-4212 (2010).
341. Liu, X.S., Li, H., Song, B. & Liu, X. Polo-like kinase 1 phosphorylation of G2 and S-phase-expressed 1 protein is essential for p53 inactivation during G2 checkpoint recovery. *EMBO Rep* **11**, 626-632 (2010).
342. Lin, Y.C., Sun, S.H. & Wang, F.F. Suppression of Polo like kinase 1 (PLK1) by p21(Waf1) mediates the p53-dependent prevention of caspase-independent mitotic death. *Cell Signal* **23**, 1816-1823 (2011).
343. Ando, K., *et al.* Polo-like kinase 1 (Plk1) inhibits p53 function by physical interaction and phosphorylation. *J Biol Chem* **279**, 25549-25561 (2004).
344. Yang, X., *et al.* Plk1-mediated phosphorylation of Topors regulates p53 stability. *J Biol Chem* **284**, 18588-18592 (2009).
345. Dias, S.S., Hogan, C., Ochocka, A.M. & Meek, D.W. Polo-like kinase-1 phosphorylates MDM2 at Ser260 and stimulates MDM2-mediated p53 turnover. *FEBS Lett* **583**, 3543-3548 (2009).
346. Bunz, F., *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497-1501 (1998).
347. Ma, S., *et al.* Induction of p21 by p65 in p53 null cells treated with Doxorubicin. *Biochim Biophys Acta* **1783**, 935-940 (2008).
348. Degenhardt, Y., *et al.* Sensitivity of cancer cells to Plk1 inhibitor GSK461364A is associated with loss of p53 function and chromosome instability. *Mol Cancer Ther* **9**, 2079-2089 (2010).
349. Gilmartin, A.G., *et al.* Distinct concentration-dependent effects of the polo-like kinase 1-specific inhibitor GSK461364A, including differential effect on apoptosis. *Cancer Res* **69**, 6969-6977 (2009).

350. Tovar, C., *et al.* Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1888-1893 (2006).
351. Vassilev, L., *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844-848 (2004).
352. Carvajal, D., *et al.* Activation of p53 by MDM2 antagonists can protect proliferating cells from mitotic inhibitors. *Cancer Res* **65**, 1918-1924 (2005).
353. Lau, L.M., Nugent, J.K., Zhao, X. & Irwin, M.S. HDM2 antagonist Nutlin-3 disrupts p73-HDM2 binding and enhances p73 function. *Oncogene* **27**, 997-1003 (2008).
354. Liu-Sullivan, N., *et al.* Pooled shRNA screen for sensitizers to inhibition of the mitotic regulator polo-like kinase (PLK1). *Oncotarget* **2**, 1254-1264 (2011).
355. Bu, Y., Yang, Z., Li, Q. & Song, F. Silencing of polo-like kinase (Plk) 1 via siRNA causes inhibition of growth and induction of apoptosis in human esophageal cancer cells. *Oncology* **74**, 198-206 (2008).
356. Zhang, Y., *et al.* Reciprocal activation between PLK1 and Stat3 contributes to survival and proliferation of esophageal cancer cells. *Gastroenterology* **142**, 521-530.e523 (2012).
357. Lin, D.C., *et al.* PLK1 Is transcriptionally activated by NF-κB during cell detachment and enhances anoikis resistance through inhibiting β-catenin degradation in esophageal squamous cell carcinoma. *Clin Cancer Res* **17**, 4285-4295 (2011).
358. Yamamoto, Y. & Gaynor, R.B. Therapeutic potential of inhibition of the NF-kappaB pathway in the treatment of inflammation and cancer. *J Clin Invest* **107**, 135-142 (2001).
359. Weiner, L.M., Surana, R. & Wang, S. Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nat Rev Immunol* **10**, 317-327 (2010).
360. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
361. Misale, S., *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532-536 (2012).
362. Wheeler, D.L., *et al.* Mechanisms of acquired resistance to cetuximab: role of HER (ErbB) family members. *Oncogene* **27**, 3944-3956 (2008).
363. Steplewski, Z., Lubeck, M.D. & Koprowski, H. Human macrophages armed with murine immunoglobulin G2a antibodies to tumors destroy human cancer cells. *Science* **221**, 865-867 (1983).
364. Dhodapkar, K.M., Krasovsky, J., Williamson, B. & Dhodapkar, M.V. Antitumor monoclonal antibodies enhance cross-presentation ofcCellular antigens and the generation of myeloma-specific killer T cells by dendritic cells. *J Exp Med* **195**, 125-133 (2002).
365. Miceli, M.C. & Parnes, J.R. The roles of CD4 and CD8 in T cell activation. *Semin Immunol* **3**, 133-141 (1991).
366. Maker, A.V., *et al.* Tumor regression and autoimmunity in patients treated with cytotoxic T lymphocyte-associated antigen 4 blockade and interleukin 2: a phase I/II study. *Ann Surg Oncol* **12**, 1005-1016 (2005).
367. Robert, C., *et al.* Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N Engl J Med* **364**, 2517-2526 (2011).
368. Rosenberg, S.A., *et al.* Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin Cancer Res* **17**, 4550-4557 (2011).
369. Morgan, R.A., *et al.* Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *J Immunother* **36**, 133-151 (2013).
370. Cameron, B.J., *et al.* Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Sci Transl Med* **5**, 197ra103 (2013).
371. Parkhurst, M.R., *et al.* T cells targeting carcinoembryonic antigen can mediate regression of metastatic colorectal cancer but induce severe transient colitis. *Mol Ther* **19**, 620-626 (2011).
372. Morgan, R.A., *et al.* Case report of a serious adverse event following the administration of T cells transduced with a chimeric antigen receptor recognizing ERBB2. *Mol Ther* **18**, 843-851 (2010).

373. Guo, K., *et al.* Targeting intracellular oncoproteins with antibody therapy or vaccination. *Sci Transl Med* **3**, 99ra85 (2011).
374. Shin, B.K., *et al.* Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J Biol Chem* **278**, 7607-7616 (2003).
375. McElroy, M., *et al.* Imaging of primary and metastatic pancreatic cancer using a fluorophore-conjugated anti-CA19-9 antibody for surgical navigation. *World J Surg* **32**, 1057-1066 (2008).
376. Wasinger, V.C., *et al.* Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **16**, 1090-1094 (1995).
377. de Godoy, L.M., *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251-1254 (2008).
378. International Human Genome Sequence Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
379. Saiki, R.K., *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491 (1988).
380. Liu, L., *et al.* Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**, 251364 (2012).
381. Pareek, C.S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *J Appl Genet* **52**, 413-435 (2011).
382. Niall, H.D. Automated Edman degradation: the protein sequenator. *Methods Enzymol* **27**, 942-1010 (1973).
383. Sperling, J., Azubel, M. & Sperling, R. Structure and function of the Pre-mRNA splicing machine. *Structure* **16**, 1605-1615 (2008).
384. Ermini, L., *et al.* Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* **18**, 1687-1693 (2008).
385. Espina, V., *et al.* A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. *Mol Cell Proteomics* **7**, 1998-2018 (2008).
386. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
387. Hu, Q., *et al.* The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430-443 (2005).
388. Hardman, M. & Makarov, A.A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Anal Chem* **75**, 1699-1705 (2003).
389. Krutchinsky, A.N., Kalkum, M. & Chait, B.T. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Anal Chem* **73**, 5066-5077 (2001).
390. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem* **59**, 2621-2625 (1987).
391. Johnson, R.S. & Taylor, J.A. Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Mol Biotechnol* **22**, 301-315 (2002).
392. Sadygov, R.G., Cociorva, D. & Yates, J.R. Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* **1**, 195-202 (2004).
393. Lam, H. Building and searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics* **10**, R111.008565 (2011).
394. Dasari, S., *et al.* Pepitome: evaluating improved spectral library search for identification complementarity and quality assessment. *J Proteome Res* **11**, 1686-1695 (2012).
395. Nesvizhskii, A.I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787-797 (2007).
396. Michalski, A., *et al.* Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol Cell Proteomics* **11**, O111.013698 (2012).
397. Jmeian, Y. & El Rassi, Z. Liquid-phase-based separation systems for depletion, prefractionation and enrichment of proteins in biological fluids for in-depth proteomics analysis. *Electrophoresis* **30**, 249-261 (2009).

398. Boisvert, F.M., Lam, Y.W., Lamont, D. & Lamond, A.I. A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol Cell Proteomics* **9**, 457-470 (2010).
399. Chenau, J., Michelland, S., Sidibe, J. & Seve, M. Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. *Proteome Sci* **6**, 9 (2008).
400. Boersema, P.J., Mohammed, S. & Heck, A.J. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal Bioanal Chem* **391**, 151-159 (2008).
401. Bantscheff, M.*, et al.* Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol Cell Proteomics* **7**, 1702-1713 (2008).
402. Garbis, S.D.*, et al.* A novel multidimensional protein identification technology approach combining protein size exclusion prefractionation, peptide zwitterion-ion hydrophilic interaction chromatography, and nano-ultraperformance RP chromatography/nESI-MS2 for the in-depth analysis of the serum proteome and phosphoproteome: application to clinical sera derived from humans with benign prostate hyperplasia. *Anal Chem* **83**, 708-718 (2011).
403. Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* **10**, 1785-1793 (2011).
404. Panchaud, A.*, et al.* Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal. Chem.* **81**, 6481-6488 (2009).
405. Scherl, A.*, et al.* Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Anal Chem* **80**, 1182-1191 (2008).
406. Panchaud, A., Jung, S., Shaffer, S.A., Aitchison, J.D. & Goodlett, D.R. Faster, quantitative, and accurate precursor acquisition independent from ion count. *Anal Chem* **83**, 2250-2257 (2011).
407. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
408. Tran, J.C.*, et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254-258 (2011).
409. Savaryn, J.P., Catherman, A.D., Thomas, P.M., Abecassis, M.M. & Kelleher, N.L. The emergence of top-down proteomics in clinical research. *Genome Med* **5**, 53 (2013).
410. Görg, A., Weiss, W. & Dunn, M.J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685 (2004).
411. Granier, F. & de Vienne, D. Silver staining of proteins: standardized procedure for two-dimensional gels bound to polyester sheets. *Anal Biochem* **155**, 45-50 (1986).
412. Dowsey, A.W., Dunn, M.J. & Yang, G.Z. The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics* **3**, 1567-1596 (2003).
413. Gerlinger, M.*, et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012).
414. Caprioli, R.M., Farmer, T.B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal Chem* **69**, 4751-4760 (1997).
415. Minerva, L., Clerens, S., Baggerman, G. & Arckens, L. Direct profiling and identification of peptide expression differences in the pancreas of control and ob/ob mice by imaging mass spectrometry. *Proteomics* **8**, 3763-3774 (2008).
416. Balluff, B., Schöne, C., Höfler, H. & Walch, A. MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications. *Histochem Cell Biol* **136**, 227-244 (2011).
417. Altelaar, A.F.*, et al.* Benchmarking stable isotope labeling based quantitative proteomics. *J Proteomics* **88**, 14-26 (2013).
418. Ong, S.E.*, et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
419. Krüger, M.*, et al.* SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353-364 (2008).

420. Sury, M.D., Chen, J.X. & Selbach, M. The SILAC fly allows for accurate protein quantification in vivo. *Mol Cell Proteomics* **9**, 2173-2183 (2010).
421. Larance, M., *et al.* Stable-isotope labeling with amino acids in nematodes. *Nat Methods* **8**, 849-851 (2011).
422. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat Methods* **7**, 383-385 (2010).
423. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**, M111.014050 (2012).
424. Gygi, S.P., *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999 (1999).
425. Zhou, H., Ranish, J.A., Watts, J.D. & Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat Biotechnol* **20**, 512-515 (2002).
426. Zhang, R., Sioma, C.S., Wang, S. & Regnier, F.E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal Chem* **73**, 5142-5149 (2001).
427. Antonov, V.K., *et al.* Studies on the mechanisms of action of proteolytic enzymes using heavy oxygen exchange. *Eur J Biochem* **117**, 195-200 (1981).
428. Hsu, J.L., Huang, S.Y., Chow, N.H. & Chen, S.H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **75**, 6843-6852 (2003).
429. Ross, P.L., *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169 (2004).
430. Thompson, A., *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
431. Bouchal, P., *et al.* Biomarker discovery in low-grade breast cancer using isobaric stable isotope tags and two-dimensional liquid chromatography-tandem mass spectrometry (iTRAQ-2DLC-MS/MS) based quantitative proteomic analysis. *J Proteome Res* **8**, 362-373 (2009).
432. Luk, V.N. & Wheeler, A.R. A digital microfluidic approach to proteomic sample processing. *Anal Chem* **81**, 4524-4530 (2009).
433. Rai, A.J., *et al.* HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* **5**, 3262-3277 (2005).
434. Lundgren, D.H., Hwang, S.I., Wu, L. & Han, D.K. Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics* **7**, 39-53 (2010).
435. Carvalho, P.C., Hewel, J., Barbosa, V.C. & Yates, J.R. Identifying differences in protein expression levels by spectral counting and feature selection. *Genet Mol Res* **7**, 342-356 (2008).
436. Griffin, N.M., *et al.* Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* **28**, 83-89 (2010).
437. Li, Z., *et al.* Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. *J Proteome Res* **11**, 1582-1590 (2012).
438. Chelius, D. & Bondarenko, P.V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J Proteome Res* **1**, 317-323 (2002).
439. Listgarten, J. & Emili, A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics* **4**, 419-434 (2005).
440. Ahmed, M.M. & Gardiner, K.J. Preserving protein profiles in tissue samples: differing outcomes with and without heat stabilization. *J Neurosci Methods* **196**, 99-106 (2011).
441. Bonner, R.F., *et al.* Laser capture microdissection: molecular analysis of tissue. *Science* **278**, 1481,1483 (1997).
442. Johann, D.J., Mukherjee, S., Prieto, D.A., Veenstra, T.D. & Blonder, J. Profiling solid tumor heterogeneity by LCM and biological MS of fresh-frozen tissue sections. *Methods Mol Biol* **755**, 95-106 (2011).
443. Wiśniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).

444. Wiśniewski, J.R., Ostasiewicz, P. & Mann, M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J Proteome Res* **10**, 3040-3049 (2011).
445. Peng, D.*, et al.* Alterations in Barrett's-related adenocarcinomas: A proteomic approach. *International Journal of Cancer* **122**, 1303-1310 (2008).
446. Langer, R.*, et al.* Protein expression profiling in esophageal adenocarcinoma patients indicates association of heat-shock protein 27 expression and chemotherapy response. *Clin Cancer Res* **14**, 8279-8287 (2008).
447. Zhao, J.*, et al.* Comparative Proteomics Analysis of Barrett Metaplasia and Esophageal Adenocarcinoma Using Two-dimensional Liquid Mass Mapping. *Molecular and Cellular Proteomics* **6**, 987-999 (2007).
448. Zhao, J.*, et al.* Comparative proteomics analysis of Barrett metaplasia and esophageal adenocarcinoma using two-dimensional liquid mass mapping. *Mol Cell Proteomics* **6**, 987-999 (2007).
449. Yoo, C.*, et al.* Automated integration of monolith-based protein separation with on-plate digestion for mass spectrometric analysis of esophageal adenocarcinoma human epithelial samples. *Electrophoresis* **27**, 3643-3651 (2006).
450. Quaas, A.*, et al.* MALDI imaging on large-scale tissue microarrays identifies molecular features associated with tumour phenotype in oesophageal cancer. *Histopathology* (2013).
451. Aichler, M.*, et al.* Clinical response to chemotherapy in oesophageal adenocarcinoma patients is linked to defects in mitochondria. *J Pathol* **230**, 410-419 (2013).
452. Elsner, M.*, et al.* MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma. *J Proteomics* **75**, 4693-4704 (2012).
453. Streitz, J.M.*, et al.* Analysis of protein expression patterns in Barrett's esophagus using MALDI mass spectrometry, in search of malignancy biomarkers. *Dis Esophagus* **18**, 170-176 (2005).
454. Farrah, T.*, et al.* The state of the human proteome in 2012 as viewed through PeptideAtlas. *J Proteome Res* **12**, 162-171 (2013).
455. Picotti, P., Bodenmiller, B. & Aebersold, R. Proteomics meets the scientific method. *Nat Methods* **10**, 24-27 (2013).
456. Janes, K.A.*, et al.* A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* **310**, 1646-1653 (2005).
457. Venkatesan, K.*, et al.* An empirical framework for binary interactome mapping. *Nat Methods* **6**, 83-90 (2009).
458. Zhu, C.*, et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res* **19**, 556-566 (2009).
459. Mo, M.L. & Palsson, B. Understanding human metabolic physiology: a genome-to-systems approach. *Trends Biotechnol* **27**, 37-44 (2009).
460. Linghu, B., Snitkin, E.S., Hu, Z., Xia, Y. & Delisi, C. Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol* **10**, R91 (2009).
461. Brohée, S. & van Helden, J. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics* **7**, 488 (2006).
462. Overton, I.M.*, et al.* Global network analysis of drug tolerance, mode of action and virulence in methicillin-resistant S. aureus. *BMC Syst Biol* **5**, 68 (2011).
463. http://www.hpacultures.org.uk/collections/ecacc.jsp. (2010)
464. http://www.sciencellonline.com/site/distributors.php. (2011)
465. Rockett, J.C., Larkin, K., Darnton, S.J., Morris, A.G. & Matthews, H.R. Five newly established oesophageal carcinoma cell lines: phenotypic and immunological characterization. *Br J Cancer* **75**, 258-263 (1997).
466. de Both, N.J., Wijnhoven, B.P., Sleddens, H.F., Tilanus, H.W. & Dinjens, W.N. Establishment of cell lines from adenocarcinomas of the esophagus and gastric cardia growing in vivo and in vitro. *Virchows Arch* **438**, 451-456 (2001).
467. Alvarez, H.*, et al.* Establishment and characterization of a bona fide Barrett esophagus-associated adenocarcinoma cell line. *Cancer Biol Ther* **7**, 1753-1755 (2008).

468. Shimada, Y., Imamura, M., Wagata, T., Yamaguchi, N. & Tobe, T. Characterization of 21 newly established esophageal cancer cell lines. *Cancer* **69**, 277-284 (1992).
469. Hasina, R., *et al.* Critical role for the receptor tyrosine kinase EPHB4 in esophageal cancers. *Cancer Res* **73**, 184-194 (2013).
470. Sehdev, V., *et al.* The aurora kinase A inhibitor MLN8237 enhances cisplatin-induced cell death in esophageal adenocarcinoma cells. *Mol Cancer Ther* **11**, 763-774 (2012).
471. Tanaka, H., *et al.* Characterization of p53 gene mutations in esophageal squamous cell carcinoma cell lines: increased frequency and different spectrum of mutations from primary tumors. *Int J Cancer* **65**, 372-376 (1996).
472. Gartler, S.M. Apparent Hela cell contamination of human heteroploid cell lines. *Nature* **217**, 750-751 (1968).
473. Lacroix, M. Persistent use of "false" cell lines. *Int J Cancer* **122**, 1-4 (2008).
474. Dirks, W.G. & Drexler, H.G. Online verification of human cell line identity by STR DNA typing. *Methods Mol Biol* **731**, 45-55 (2011).
475. Dirks, W.G. & Drexler, H.G. STR DNA typing of human cell lines: detection of intra- and interspecies cross-contamination. *Methods Mol Biol* **946**, 27-38 (2013).
476. Masters, J.R., *et al.* Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc Natl Acad Sci U S A* **98**, 8012-8017 (2001).
477. Budowle, B., Shea, B., Niezgoda, S. & Chakraborty, R. CODIS STR loci data from 41 sample populations. *J Forensic Sci* **46**, 453-489 (2001).
478. Sambrook, J. & Russell, D.W. *Molecular cloning : a laboratory manual*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001).
479. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254 (1976).
480. Mathews, S.T., Plaisance, E.P. & Kim, T. Imaging systems for westerns: chemiluminescence vs. infrared detection. *Methods Mol Biol* **536**, 499-513 (2009).
481. Wang, Y.V., *et al.* Quantitative analyses reveal the importance of regulated Hdmx degradation for p53 activation. *Proc Natl Acad Sci U S A* **104**, 12365-12370 (2007).
482. Gerk, P.M. Quantitative immunofluorescent blotting of the multidrug resistance-associated protein 2 (MRP2). *J Pharmacol Toxicol Methods* **63**, 279-282 (2011).
483. Bond, D., Primrose, D.A. & Foley, E. Quantitative evaluation of signaling events in Drosophila S2 cells. *Biol Proced Online* **10**, 20-28 (2008).
484. Sowell, J., Strekowski, L. & Patonay, G. DNA and protein applications of near-infrared dyes. *J Biomed Opt* **7**, 571-575 (2002).
485. Riss, T.L., Moravec, R.A. & Niles, A.L. Cytotoxicity testing: measuring viable cells, dead cells, and detecting mechanism of cell death. *Methods Mol Biol* **740**, 103-114 (2011).
486. O'Brien, J., Wilson, I., Orton, T. & Pognan, F. Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur J Biochem* **267**, 5421-5426 (2000).
487. Zhang, J.H., Chung, T.D. & Oldenburg, K.R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67-73 (1999).
488. O'Neill, M., Campbell, S.J., Save, V., Thompson, A.M. & Hall, P.A. An immunochemical analysis of mdm2 expression in human breast cancer and the identification of a growth-regulated cross-reacting species p170. *J Pathol* **186**, 254-261 (1998).
489. Limame, R., *et al.* Comparative analysis of dynamic cell viability, migration and invasion assessments by novel real-time technology and classic endpoint assays. *PLoS One* **7**, e46536 (2012).
490. Bonnet, G., Tyagi, S., Libchaber, A. & Kramer, F.R. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc Natl Acad Sci U S A* **96**, 6171-6176 (1999).
491. Ogrean, C., Jackson, B. & Covino, J. Quantitative real-time PCR using the thermo scientific Solaris qPCR assay. *J Vis Exp* (2010).
492. Gibson, U.E., Heid, C.A. & Williams, P.M. A novel method for real time quantitative RT-PCR. *Genome Res* **6**, 995-1001 (1996).

493. Bustin, S.A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**, 169-193 (2000).
494. Vandesompele, J.*, et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, (2002).
495. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).
496. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).
497. Rubie, C.*, et al.* Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol Cell Probes* **19**, 101-109 (2005).
498. Taylor, J.R. *An introduction to error analysis : the study of uncertainties in physical measurements*, (University Science Books, Sausalito, Calif., 1997).
499. Allred, D.C., Harvey, J.M., Berardo, M. & Clark, G.M. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol* **11**, 155-168 (1998).
500. http://icgc.org/icgc/cgp/72/508/70708. (2012).
501. Viera, A.J. & Garrett, J.M. Understanding interobserver agreement: the kappa statistic. *Fam Med* **37**, 360-363 (2005).
502. Lowry, O., Rosebrough, N., Farr, A. & Randall, R. Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**, 265-275 (1951).
503. Hörth, P., Miller, C.A., Preckel, T. & Wenz, C. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol Cell Proteomics* **5**, 1968-1974 (2006).
504. Gluck, F.*, et al.* EasyProt--an easy-to-use graphical platform for proteomics data analysis. *J Proteomics* **79**, 146-160 (2013).
505. Dayon, L., Pasquarello, C., Hoogland, C., Sanchez, J.C. & Scherl, A. Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. *J Proteomics* **73**, 769-777 (2010).
506. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
507. JP, S. Caution on the Use of Variance Ratio: A Comment. *Review of Educational Research* **62**, 429-432 (1992).
508. Borenstein, M. *Introduction to meta-analysis*, (John Wiley & Sons, Chichester, U.K., 2009).
509. Shapiro, S. & Wilk, M. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591-611 (1965).
510. Lilliefors, H. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* **62**, 399-402 (1967).
511. Conover, W., Johnson, M. & Johnson, M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23**, 351-361 (1981).
512. Ruxton, G. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioural Ecology* **17**, 688-690 (2006).
513. Welch, B. The generalization of "Student's" problem when several different population variances are involved. *Biometrika* **34**, 28-35 (1947).
514. Satterthwaite, F.E. An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2**, 110-114 (1946).
515. Benjamini, Y. & Yekutiele, D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165-1185 (2001).
516. Apweiler, R.*, et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115-119 (2004).
517. Joshua, A.M. & Boutros, P.C. Web-based resources for clinical bioinformatics. *Methods Mol Med* **141**, 309-329 (2008).
518. Roberts, R.J. PubMed Central: The GenBank of the published literature. *Proc Natl Acad Sci U S A* **98**, 381-382 (2001).
519. Rhodes, D.R.*, et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1-6 (2004).
520. Brazma A & J, V. Gene expression data analysis. *FEBS Letters* **480**, 17-24 (2000).

521. Yates, J.R., Eng, J.K., McCormack, A.L. & Schieltz, D. Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* **67**, 1426-1436 (1995).
522. Deutsch, E.W., *et al.* A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10**, 1150-1159 (2010).
523. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).
524. Shannon, P., *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
525. Huang, d.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
526. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
527. Kanehisa, M. Molecular network analysis of diseases and drugs in KEGG. *Methods Mol Biol* **939**, 263-275 (2013).
528. Szklarczyk, D., *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39**, D561-568 (2011).
529. Brown, C.J., Lain, S., Verma, C.S., Fersht, A.R. & Lane, D.P. Awakening guardian angels: drugging the p53 pathway. *Nat Rev Cancer* **9**, 862-873 (2009).
530. Ohkubo, S., Tanaka, T., Taya, Y., Kitazato, K. & Prives, C. Excess HDM2 impacts cell cycle and apoptosis and has a selective effect on p53-dependent transcription. *J Biol Chem* **281**, 16943-16950 (2006).
531. Wang, Y.V., *et al.* Quantitative analyses reveal the importance of regulated Hdmx degradation for p53 activation. *Proc Natl Acad Sci U S A* **104**, 12365-12370 (2007).
532. http://www.lgcstandards-atcc.org/Products/Cells_and_Microorganisms/hTERT_Immortalized_Cell_Lines/Barretts_Esophageal_Epithelial_Cells/CRL-4027.aspx#A7931A04156C4C7FA40828AEF707302F. (2010)
533. Jackson, A.L., *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**, 635-637 (2003).
534. Fedorov, Y., *et al.* Off-target effects by siRNA can induce toxic phenotype. *RNA* **12**, 1188-1196 (2006).
535. Vojtěsek, B., Bártek, J., Midgley, C.A. & Lane, D.P. An immunochemical analysis of the human nuclear phosphoprotein p53. New monoclonal antibodies and epitope mapping using recombinant p53. *J Immunol Methods* **151**, 237-244 (1992).
536. Vojtěsek, B., Fisher, C.J., Barnes, D.M. & Lane, D.P. Comparison between p53 staining in tissue sections and p53 proteins levels measured by an ELISA technique. *Br J Cancer* **67**, 1254-1258 (1993).
537. Larsen, H., Roug, A.S., Nielsen, K., Søndergaard, C.S. & Hokland, P. Nonviral transfection of leukemic primary cells and cells lines by siRNA-a direct comparison between Nucleofection and Accell delivery. *Exp Hematol* **39**, 1081-1089 (2011).
538. Lowery, D.M., *et al.* Proteomic screen defines the Polo-box domain interactome and identifies Rock2 as a Plk1 substrate. *EMBO J* **26**, 2262-2273 (2007).
539. Hollick, J.J., Jones, S.D., Flynn, C.J. & Thomas, M.G. Pyrimidine Derivatives as Protein Kinase Inhibitors. *US Patent Office Application* 1-178 (Cyclacel Limited, 2008)
540. Frame, S., *et al.* Potent and selective small molecule inhibitors of Polo-like kinase 1: biological characterization. in *Proceedings of the 103rd Annual Meeting of the American Association of Cancer Research* (Chicago, Illinois, 2012).
541. Obach, R.S., *et al.* The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J Pharmacol Exp Ther* **283**, 46-58 (1997).
542. van Hennik, M.B., *et al.* Comparative pharmacokinetics of cisplatin and three analogues in mice and humans. *Cancer Res* **47**, 6297-6301 (1987).
543. Krishan, A. Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *J Cell Biol* **66**, 188-193 (1975).
544. Sorenson, C.M. & Eastman, A. Mechanism of cis-diamminedichloroplatinum(II)-induced cytotoxicity: role of G2 arrest and DNA double-strand breaks. *Cancer Res* **48**, 4484-4488 (1988).

545. Sancho-Martínez, S.M., Prieto-García, L., Prieto, M., López-Novoa, J.M. & López-Hernández, F.J. Subcellular targets of cisplatin cytotoxicity: an integrated view. *Pharmacol Ther* **136**, 35-55 (2012).
546. McMillin, D.W., *et al.* Microenvironmental influence on pre-clinical activity of polo-like kinase inhibition in multiple myeloma: implications for clinical translation. *PLoS One* **6**, e20226 (2011).
547. Cheok, C., Kua, N., Kaldis, P. & Lane, D. Combination of nutlin-3 and VX-680 selectively targets p53 mutant cells with reversible effects on cells expressing wild-type p53. *Cell Death Differ* **17**, 1486-1500 (2010).
548. Huang, B., Deo, D., Xia, M. & Vassilev, L.T. Pharmacologic p53 activation blocks cell cycle progression but fails to induce senescence in epithelial cancer cells. *Mol Cancer Res* **7**, 1497-1509 (2009).
549. Komlodi-Pasztor, E., Sackett, D., Wilkerson, J. & Fojo, T. Mitosis is not a key target of microtubule agents in patient tumors. *Nat Rev Clin Oncol* **8**, 244-250 (2011).
550. Dulak, A.M., *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* (2013).
551. Milner, J., Medcalf, E.A. & Cook, A.C. Tumor suppressor p53: analysis of wild-type and mutant p53 complexes. *Mol Cell Biol* **11**, 12-19 (1991).
552. García-Cao, I., *et al.* "Super p53" mice exhibit enhanced DNA damage response, are tumor resistant and age normally. *EMBO J* **21**, 6225-6235 (2002).
553. Tyner, S.D., *et al.* p53 mutant mice that display early ageing-associated phenotypes. *Nature* **415**, 45-53 (2002).
554. Chen, H., *et al.* A cell-based immunocytochemical assay for monitoring kinase signaling pathways and drug efficacy. *Anal Biochem* **338**, 136-142 (2005).
555. Suzuki, K. & Takahashi, K. Reduced cell adhesion during mitosis by threonine phosphorylation of beta1 integrin. *J Cell Physiol* **197**, 297-305 (2003).
556. Chène, P. & Bechter, E. p53 mutants without a functional tetramerisation domain are not oncogenic. *J Mol Biol* **286**, 1269-1274 (1999).
557. Ambrosini, G., *et al.* Mouse double minute antagonist Nutlin-3a enhances chemotherapy-induced apoptosis in cancer cells with mutant p53 by activating E2F1. *Oncogene* **26**, 3473-3481 (2007).
558. Ray, R.M., Bhattacharya, S. & Johnson, L.R. Mdm2 inhibition induces apoptosis in p53 deficient human colon cancer cells by activating p73- and E2F1-mediated expression of PUMA and Siva-1. *Apoptosis* **16**, 35-44 (2011).
559. Nicholson, J., *et al.* An iTRAQ proteomics screen reveals the effects of the MDM2 binding ligand Nutlin-3 on cellular proteostasis. *J Proteome Res* **11**, 5464-5478 (2012).
560. Hall, P.A. Assessing apoptosis: a critical survey. *Endocr Relat Cancer* **6**, 3-8 (1999).
561. Samuel, M.A., Morrey, J.D. & Diamond, M.S. Caspase 3-dependent cell death of neurons contributes to the pathogenesis of West Nile virus encephalitis. *J Virol* **81**, 2614-2623 (2007).
562. Zhang, J., *et al.* Visualization of caspase-3-like activity in cells using a genetically encoded fluorescent biosensor activated by protein cleavage. *Nat Commun* **4**, 2157 (2013).
563. Keppner, S., Proschak, E., Schneider, G. & Spänkuch, B. Fate of primary cells at the G_1/S boundary after polo-like kinase 1 inhibition by SBE13. *Cell Cycle* **10**, 708-720 (2011).
564. Gündisch, S., *et al.* Variability of protein and phosphoprotein levels in clinical tissue specimens during the preanalytical phase. *J Proteome Res* **11**, 5748-5762 (2012).
565. Geiser, L., Dayon, L., Vaezzadeh, A.R. & Hochstrasser, D.F. Shotgun proteomics: a relative quantitative approach using Off-Gel electrophoresis and LC-MS/MS. *Methods Mol Biol* **681**, 459-472 (2011).
566. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29**, 1165-1185 (2001).
567. Wlazlinski, A., *et al.* Downregulation of several fibulin genes in prostate cancer. *Prostate* **67**, 1770-1780 (2007).
568. Cheng, Y.Y., *et al.* Fibulin 1 is downregulated through promoter hypermethylation in gastric cancer. *Br J Cancer* **99**, 2083-2087 (2008).
569. Kanda, M., *et al.* Promoter hypermethylation of fibulin 1 gene is associated with tumor progression in hepatocellular carcinoma. *Mol Carcinog* **50**, 571-579 (2011).

570. Bao, W., *et al.* Silencing of Cathepsin B suppresses the proliferation and invasion of endometrial cancer. *Oncol Rep* **30**, 723-730 (2013).
571. Yin, M., *et al.* TGF-β signaling, activated stromal fibroblasts, and cysteine cathepsins B and L drive the invasive growth of human melanoma cells. *Am J Pathol* **181**, 2202-2216 (2012).
572. Withana, N.P., *et al.* Cathepsin B inhibition limits bone metastasis in breast cancer. *Cancer Res* **72**, 1199-1209 (2012).
573. Ding, S., *et al.* Molecular imaging of gastric neoplasia with near-infrared fluorescent activatable probes. *Mol Imaging* **11**, 507-515 (2012).
574. Gologan, A., Acquafondata, M., Dhir, R. & Sepulveda, A.R. Polymeric immunoglobulin receptor-negative tumors represent a more aggressive type of adenocarcinomas of distal esophagus and gastroesophageal junction. *Arch Pathol Lab Med* **132**, 1295-1301 (2008).
575. Dumartin, L., *et al.* AGR2 is a novel surface antigen that promotes the dissemination of pancreatic cancer cells through regulation of cathepsins B and D. *Cancer Res* **71**, 7091-7102 (2011).
576. Chen, M., *et al.* Promoter hypermethylation mediated downregulation of FBP1 in human hepatocellular carcinoma and colon cancer. *PLoS One* **6**, e25564 (2011).
577. Liu, X., *et al.* Warburg effect revisited: an epigenetic link between glycolysis and gastric carcinogenesis. *Oncogene* **29**, 442-450 (2010).
578. Nalesnik, M.A., *et al.* Gene deletions and amplifications in human hepatocellular carcinomas: correlation with hepatocyte growth regulation. *Am J Pathol* **180**, 1495-1508 (2012).
579. Jeong, H.C., *et al.* Proteomic analysis of human small cell lung cancer tissues: up-regulation of coactosin-like protein-1. *J Proteome Res* **10**, 269-276 (2011).
580. Hassan, M.I., Toor, A. & Ahmad, F. Progastriscin: structure, function, and its role in tumor progression. *J Mol Cell Biol* **2**, 118-127 (2010).
581. Ernst, M., *et al.* STAT3 and STAT1 mediate IL-11-dependent and inflammation-associated gastric tumorigenesis in gp130 receptor mutant mice. *J Clin Invest* **118**, 1727-1738 (2008).
582. Breton, J., *et al.* Proteomic screening of a cell line model of esophageal carcinogenesis identifies cathepsin D and aldo-keto reductase 1C2 and 1B10 dysregulation in Barrett's esophagus and esophageal adenocarcinoma. *J Proteome Res* **7**, 1953-1962 (2008).
583. Klopfleisch, R., *et al.* Proteome of metastatic canine mammary carcinomas: similarities to and differences from human breast cancer. *J Proteome Res* **9**, 6380-6391 (2010).
584. Alvarez, H., *et al.* Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet* **7**, e1001356 (2011).
585. Fareed, K.R., *et al.* Caspase-cleaved cytokeratin-18 and tumour regression in gastro-oesophageal adenocarcinomas treated with neoadjuvant chemotherapy. *World J Gastroenterol* **18**, 1915-1920 (2012).
586. Langer, R., *et al.* Comparison of pretherapeutic and posttherapeutic expression levels of chemotherapy-associated genes in adenocarcinomas of the esophagus treated by 5-fluorouracil- and cisplatin-based neoadjuvant chemotherapy. *Am J Clin Pathol* **128**, 191-197 (2007).
587. Mushinski, J.F., *et al.* Inhibition of tumor cell motility by the interferon-inducible GTPase MxA. *J Biol Chem* **284**, 15206-15214 (2009).
588. Gillen, P., McDermott, M., Grehan, D., Hourihane, D.O. & Hennessy, T.P. Proliferating cell nuclear antigen in the assessment of Barrett's mucosa. *Br J Surg* **81**, 1766-1768 (1994).
589. Lagarde, S.M., *et al.* Analysis of gene expression identifies differentially expressed genes and pathways associated with lymphatic dissemination in patients with adenocarcinoma of the esophagus. *Ann Surg Oncol* **15**, 3459-3470 (2008).
590. Kang, R., Zhang, Q., Zeh, H.J., Lotze, M. & Tang, D. HMGB1 in Cancer: Good, Bad, or Both? *Clin Cancer Res* (2013).
591. Ren, H., *et al.* Analysis of variabilities of serum proteomic spectra in patients with gastric cancer before and after operation. *World J Gastroenterol* **12**, 2789-2792 (2006).
592. Ryu, J.W., *et al.* The proteomics approach to find biomarkers in gastric cancer. *J Korean Med Sci* **18**, 505-509 (2003).

593. Kang, S., *et al.* Molecular proteomics imaging of tumor interfaces by mass spectrometry. *J Proteome Res* **9**, 1157-1164 (2010).
594. Langer, R., Feith, M., Siewert, J.R., Wester, H.J. & Hoefler, H. Expression and clinical significance of glucose regulated proteins GRP78 (BiP) and GRP94 (GP96) in human adenocarcinomas of the esophagus. *BMC Cancer* **8**, 70 (2008).
595. Altorjay, A., *et al.* Significance and prognostic value of lysosomal enzyme activities measured in surgically operated adenocarcinomas of the gastroesophageal junction and squamous cell carcinomas of the lower third of esophagus. *World J Gastroenterol* **11**, 5751-5756 (2005).
596. Hughes, S.J., *et al.* A novel amplicon at 8p22-23 results in overexpression of cathepsin B in esophageal adenocarcinoma. *Proc Natl Acad Sci U S A* **95**, 12410-12415 (1998).
597. Golan-Gerstl, R., *et al.* Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res* **71**, 4464-4472 (2011).
598. Wang, X.M., *et al.* PTP1B contributes to calreticulin-induced metastatic phenotypes in esophageal squamous cell carcinoma. *Mol Cancer Res* **11**, 986-994 (2013).
599. Du, X.L., *et al.* Calreticulin promotes cell motility and enhances resistance to anoikis through STAT3-CTTN-Akt pathway in esophageal squamous cell carcinoma. *Oncogene* **28**, 3714-3722 (2009).
600. Petrova, D.T., *et al.* Expression of chloride intracellular channel protein 1 (CLIC1) and tumor protein D52 (TPD52) as potential biomarkers for colorectal cancer. *Clin Biochem* **41**, 1224-1236 (2008).
601. Piaggi, S., *et al.* Nuclear translocation of glutathione transferase omega is a progression marker in Barrett's esophagus. *Oncol Rep* **21**, 283-287 (2009).
602. Tsuboi, K., *et al.* Potent and selective inhibitors of glutathione S-transferase omega 1 that impair cancer drug resistance. *J Am Chem Soc* **133**, 16605-16616 (2011).
603. Li, S.L., *et al.* Quantitative proteome analysis of multidrug resistance in human ovarian cancer cell line. *J Cell Biochem* **109**, 625-633 (2010).
604. Yamaguchi, H. & Condeelis, J. Regulation of the actin cytoskeleton in cancer cell migration and invasion. *Biochim Biophys Acta* **1773**, 642-652 (2007).
605. Darragh, J., *et al.* The calcium-binding domain of the stress protein SEP53 is required for survival in response to deoxycholic acid-mediated injury. *FEBS J* **273**, 1930-1947 (2006).
606. Opitz, O.G., Jenkins, T.D. & Rustgi, A.K. Transcriptional regulation of the differentiation-linked human K4 promoter is dependent upon esophageal-specific nuclear factors. *J Biol Chem* **273**, 23912-23921 (1998).
607. Fox, C.A., *et al.* Altered Expression of TFF-1 and CES-2 in Barrett's Esophagus and Associated Adenocarcinoma. *Neoplasia* **7**, 407-416 (2005).
608. Luthra, R., *et al.* Gene Expression Profiling of Localised Esophageal Carcinomas: Association with Pathological response to Preoperative Chemoradiation. *Journal of Clinical Oncology* **24**, 259-267 (2006).
609. Lee, J.H., Jang, S.I., Yang, J.M., Markova, N.G. & Steinert, P.M. The proximal promoter of the human transglutaminase 3 gene. Stratified squamous epithelial-specific expression in cultured cells is mediated by binding of Sp1 and ets transcription factors to a proximal promoter element. *J Biol Chem* **271**, 4561-4568 (1996).
610. Roesch-Ely, M., Schnölzer, M., Nees, M., Plinkert, P.K. & Bosch, F.X. Reference spectra from squamous epithelium and connective tissue allow whole section proteomics analysis. *Arch Physiol Biochem* **116**, 218-226 (2010).
611. Dong, C., *et al.* Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. *Cancer Cell* **23**, 316-331 (2013).
612. Ramos-Vara, J.A. Principles and methods of immunohistochemistry. *Methods Mol Biol* **691**, 83-96 (2011).
613. Pohler, E., *et al.* The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. *Mol Cell Proteomics* **3**, 534-547 (2004).
614. Pizzi, M., *et al.* Anterior gradient 2 profiling in Barrett columnar epithelia and adenocarcinoma. *Hum Pathol* **43**, 1839-1844 (2012).
615. Thompson, D.A. & Weigel, R.J. hAG-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. *Biochem Biophys Res Commun* **251**, 111-116 (1998).

616. Uhlén, M., *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **4**, 1920-1932 (2005).
617. Lee, d.H., *et al.* Identification of proteins differentially expressed in gastric cancer cells with high metastatic potential for invasion to lymph nodes. *Mol Cells* **31**, 563-571 (2011).
618. Baldauf, H.M., *et al.* SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat Med* **18**, 1682-1687 (2012).
619. Hetz, C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat Rev Mol Cell Biol* **13**, 89-102 (2012).
620. Paton, A.W., *et al.* AB5 subtilase cytotoxin inactivates the endoplasmic reticulum chaperone BiP. *Nature* **443**, 548-552 (2006).
621. Scherle, P., Behrens, T. & Staudt, L.M. Ly-GDI, a GDP-dissociation inhibitor of the RhoA GTP-binding protein, is expressed preferentially in lymphocytes. *Proc Natl Acad Sci U S A* **90**, 7568-7572 (1993).
622. Moissoglu, K., McRoberts, K.S., Meier, J.A., Theodorescu, D. & Schwartz, M.A. Rho GDP dissociation inhibitor 2 suppresses metastasis via unconventional regulation of RhoGTPases. *Cancer Res* **69**, 2838-2844 (2009).
623. Cho, H.J., *et al.* RhoGDI2 expression is associated with tumor growth and malignant progression of gastric cancer. *Clin Cancer Res* **15**, 2612-2619 (2009).
624. Abnova. http://www.abnova.com/products/products_detail.asp?Catalog_id=H00000397-A01. (Accessed May 25th 2013.).
625. Wei, L., *et al.* Hsp27 participates in the maintenance of breast cancer stem cells through regulation of epithelial-mesenchymal transition and nuclear factor-κB. *Breast Cancer Res* **13**, R101 (2011).
626. Vargha, R., *et al.* Effects of epithelial-to-mesenchymal transition on acute stress response in human peritoneal mesothelial cells. *Nephrol Dial Transplant* **23**, 3494-3500 (2008).
627. Mizutani, H., *et al.* HSP27 modulates epithelial to mesenchymal transition of lung cancer cells in a Smad-independent manner. *Oncol Lett* **1**, 1011-1016 (2010).
628. Shiota, M., *et al.* Hsp27 regulates epithelial mesenchymal transition, metastasis, and circulating tumor cells in prostate cancer. *Cancer Res* **73**, 3109-3119 (2013).
629. Soldes, O.S., *et al.* Differential expression of Hsp27 in normal oesophagus, Barrett's metaplasia and oesophageal adenocarcinomas. *Br J Cancer* **79**, 595-603 (1999).
630. Doak, S.H., *et al.* Differential expression of the MAD2, BUB1 and HSP27 genes in Barrett's oesophagus-their association with aneuploidy and neoplastic progression. *Mutat Res* **547**, 133-144 (2004).
631. Lambot, M.A., Peny, M.O., Fayt, I., Haot, J. & Noël, J.C. Overexpression of 27-kDa heat shock protein relates to poor histological differentiation in human oesophageal squamous cell carcinoma. *Histopathology* **36**, 326-330 (2000).
632. Slotta-Huspenina, J., *et al.* Evidence of prognostic relevant expression profiles of heat-shock proteins and glucose-regulated proteins in oesophageal adenocarcinomas. *PLoS One* **7**, e41420 (2012).
633. Wiśniewski, J.R., Zielinska, D.F. & Mann, M. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Anal Biochem* **410**, 307-309 (2011).
634. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).
635. McGill, R., Tukey, J. & Larsen, W. Variations of box plots. *The American Statistician* **32**, 12-16 (1978).
636. Bayes, T. & Price, R. An Essay towards solving a Problem in the Doctrine of Chance. *Philosophical Transactions of the Royal Society of London* **53**, 370-418 (1763).
637. Jaynes, E. *Probability Theory: The Logic of Science*, (Cambridge University Press, Cambridge, 2003).
638. Aickin, M. Bayes without priors. *J Clin Epidemiol* **57**, 4-13 (2004).
639. Acosta-Martin, A.E., *et al.* Quantitative mass spectrometry analysis using PAcIFIC for the identification of plasma diagnostic biomarkers for abdominal aortic aneurysm. *PLoS One* **6**, e28698 (2011).

640. Wiśniewski, J.R., *et al.* Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol Syst Biol* **8**, 611 (2012).
641. Malakoff, D. Bayes offers a 'new' way to make sense of numbers. *Science* **286**, 1460-1464 (1999).
642. Wu, S.H., Black, M.A., North, R.A., Atkinson, K.R. & Rodrigo, A.G. A statistical model to identify differentially expressed proteins in 2D PAGE gels. *PLoS Comput Biol* **5**, e1000509 (2009).
643. Krogh, M., Fernandez, C., Teilum, M., Bengtsson, S. & James, P. A probabilistic treatment of the missing spot problem in 2D gel electrophoresis experiments. *J Proteome Res* **6**, 3335-3343 (2007).
644. Regalado, S.P., Nambu, Y., Iannettoni, M.D., Orringer, M.B. & Beer, D.G. Abundant expression of the intestinal protein villin in Barrett's metaplasia and esophageal adenocarcinomas. *Mol Carcinog* **22**, 182-189 (1998).
645. Iannettoni, M.D., *et al.* Detection of Barrett's adenocarcinoma of the gastric cardia with sucrase isomaltase and p53. *Ann Thorac Surg* **62**, 1460-1465; discussion 1465-1466 (1996).
646. Subramanian, A., *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
647. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18 Suppl 1**, S233-240 (2002).
648. Liu, M., *et al.* Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genet* **3**, e96 (2007).
649. Mishra, G.R., *et al.* Human protein reference database - 2006 update. *Nucleic Acids Res* **34**, D411-414 (2006).
650. Bader, G.D., Betel, D. & Hogue, C.W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res* **31**, 248-250 (2003).
651. Breitkreutz, B.J., *et al.* The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res* **36**, D637-640 (2008).
652. Kerrien, S., *et al.* IntAct--open source resource for molecular interaction data. *Nucleic Acids Res* **35**, D561-565 (2007).
653. Mewes, H.W., *et al.* MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res* **36**, D196-201 (2008).
654. Salwinski, L., *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* **32**, D449-451 (2004).
655. Chatr-Aryamontri, A., Zanzoni, A., Ceol, A. & Cesareni, G. Searching the protein interaction space through the MINT database. *Methods Mol Biol* **484**, 305-317 (2008).
656. Rual, J.F., *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-1178 (2005).
657. Ewing, R.M., *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**, 89 (2007).
658. Rhodes, D.R., *et al.* Probabilistic model of the human protein-protein interaction network. *Nat Biotechnol* **23**, 951-959 (2005).
659. Scott, M.S. & Barton, G.J. Probabilistic prediction and ranking of human protein-protein interactions. *BMC Bioinformatics* **8**, 239 (2007).
660. von Mering, C., *et al.* STRING 7- recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**, D358-362 (2007).
661. Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J. & Pavlidis, P. Coexpression analysis of human genes across many microarray data sets. *Genome Res* **14**, 1085-1094 (2004).
662. Jensen, L.J., Lagarde, J., von Mering, C. & Bork, P. ArrayProspector: a web resource of functional associations inferred from microarray expression data. *Nucleic Acids Res* **32**, W445-448 (2004).
663. Griffith, O.L., *et al.* Assessment and integration of publicly available SAGE, cDNA microarray, and oligonucleotide microarray expression data for global coexpression analyses. *Genomics* **86**, 476-488 (2005).

664. Mulder, N.J., *et al.* InterPro, progress and status in 2005. *Nucleic Acids Res* **33**, D201-205 (2005).
665. Lee, I., Li, Z. & Marcotte, E.M. An improved, bias-reduced probabilistic functional gene network of baker's yeast, Saccharomyces cerevisiae. *PLoS One* **2**, e988 (2007).
666. Lee, I., *et al.* A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans. *Nat Genet* **40**, 181-188 (2008).
667. Harris, M.A., *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**, D258-261 (2004).
668. van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks. *Methods Mol Biol* **804**, 281-295 (2012).
669. Dennis, G., *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
670. Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* **11**, R53 (2010).
671. Peri, S., *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* **13**, 2363-2371 (2003).
672. Finn, R.D., *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247-251 (2006).
673. Ashburner, M., *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
674. Rzhetsky, A., *et al.* GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J Biomed Inform* **37**, 43-53 (2004).
675. Solinas, G., Marchesi, F., Garlanda, C., Mantovani, A. & Allavena, P. Inflammation-mediated promotion of invasion and metastasis. *Cancer Metastasis Rev* **29**, 243-248 (2010).
676. Qian, B., *et al.* A distinct macrophage population mediates metastatic breast cancer cell extravasation, establishment and growth. *PLoS One* **4**, e6562 (2009).
677. Joyce, J.A. & Pollard, J.W. Microenvironmental regulation of metastasis. *Nat Rev Cancer* **9**, 239-252 (2009).
678. Condeelis, J. & Pollard, J.W. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell* **124**, 263-266 (2006).
679. Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. *Nature* **454**, 436-444 (2008).
680. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029-1033 (2009).
681. Menon, R., *et al.* Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer. *Cancer Res* **69**, 300-309 (2009).
682. Kalyana-Sundaram, S., *et al.* Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **149**, 1622-1634 (2012).
683. Geiger, T., *et al.* Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Mol Cell Proteomics* **12**, 1709-1722 (2013).
684. Lantz, B. The impact of sample non-normality on ANOVA and alternative methods. *Br J Math Stat Psychol* **66**, 224-244 (2013).
685. Vehviläinen, P., Hyytiäinen, M. & Keski-Oja, J. Latent transforming growth factor-beta-binding protein 2 is an adhesion protein for melanoma cells. *J Biol Chem* **278**, 24705-24713 (2003).
686. Kosanam, H., *et al.* LAMC2: A promising new pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues. *Mol Cell Proteomics* (2013).
687. Warburg, O. On the origin of cancer cells. *Science* **123**, 309-314 (1956).
688. von Mering, C., *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* **33**, D433-437 (2005).
689. Anders, M., *et al.* Expression of EpCam and villin in Barrett's esophagus and in gastric cardia. *Dis Markers* **24**, 287-292 (2008).
690. Seimetz, D., Lindhofer, H. & Bokemeyer, C. Development and approval of the trifunctional antibody catumaxomab (anti-EpCAM x anti-CD3) as a targeted cancer immunotherapy. *Cancer Treat Rev* **36**, 458-467 (2010).

691. Sakamoto, J., *et al.* A phase I radioimmunolocalization trial of humanized monoclonal antibody huA33 in patients with gastric carcinoma. *Cancer Sci* **97**, 1248-1254 (2006).
692. Chong, G., *et al.* Phase I trial of 131I-huA33 in patients with advanced colorectal carcinoma. *Clin Cancer Res* **11**, 4818-4826 (2005).
693. Vansteenkiste, J., *et al.* Adjuvant MAGE-A3 Immunotherapy in Resected Non-Small-Cell Lung Cancer: Phase II Randomized Study Results. *J Clin Oncol* **31**, 2396-2403 (2013).
694. Theocharidis, A., van Dongen, S., Enright, A.J. & Freeman, T.C. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nat Protoc* **4**, 1535-1550 (2009).
695. Wu, C., Macleod, I. & Su, A.I. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res* **41**, D561-565 (2013).
696. Venables, J.P., *et al.* Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16**, 670-676 (2009).
697. Freeman, T.C., *et al.* A gene expression atlas of the domestic pig. *BMC Biol* **10**, 90 (2012).
698. Goebel, M., Stengel, A., Lambrecht, N.W. & Sachs, G. Selective gene expression by rat gastric corpus epithelium. *Physiol Genomics* **43**, 237-254 (2011).
699. Estrella, V., *et al.* Acidity generated by the tumor microenvironment drives local invasion. *Cancer Res* **73**, 1524-1535 (2013).
700. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**, 504-507 (2001).
701. Ott, K., Weber, W. & Siewert, J.R. The importance of PET in the diagnosis and response evaluation of esophageal cancer. *Dis Esophagus* **19**, 433-442 (2006).
702. Younes, M., *et al.* Relationship between dysplasia, p53 protein accumulation, DNA ploidy, and Glut1 overexpression in Barrett metaplasia. *Scand J Gastroenterol* **35**, 131-137 (2000).
703. Fonteyne, P., *et al.* Expression of hexokinases and glucose transporters in treated and untreated oesophageal adenocarcinoma. *Histol Histopathol* **24**, 971-977 (2009).
704. Liu, Y., *et al.* A small-molecule inhibitor of glucose transporter 1 downregulates glycolysis, induces cell-cycle arrest, and inhibits cancer cell growth in vitro and in vivo. *Mol Cancer Ther* **11**, 1672-1682 (2012).
705. Rastogi, S., Banerjee, S., Chellappan, S. & Simon, G.R. Glut-1 antibodies induce growth arrest and apoptosis in human cancer cell lines. *Cancer Lett* **257**, 244-251 (2007).
706. Chan, D.A., *et al.* Targeting GLUT1 and the Warburg effect in renal cell carcinoma by chemical synthetic lethality. *Sci Transl Med* **3**, 94ra70 (2011).
707. Christofk, H.R., *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230-233 (2008).
708. Bluemlein, K., *et al.* No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis. *Oncotarget* **2**, 393-400 (2011).
709. Mazurek, S. Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells. *Int J Biochem Cell Biol* **43**, 969-980 (2011).
710. Morgan, H.P., *et al.* M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation. *Proc Natl Acad Sci U S A* **110**, 5881-5886 (2013).
711. Anastasiou, D., *et al.* Pyruvate kinase M2 activators promote tetramer formation and suppress tumorigenesis. *Nat Chem Biol* **8**, 839-847 (2012).
712. Beyenbach, K.W. & Wieczorek, H. The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. *J Exp Biol* **209**, 577-589 (2006).
713. Wiedmann, R.M., *et al.* The V-ATPase-inhibitor archazolid abrogates tumor metastasis via inhibition of endocytic activation of the Rho-GTPase Rac1. *Cancer Res* **72**, 5976-5987 (2012).
714. von Schwarzenberg, K., *et al.* Mode of cell death induction by pharmacological vacuolar H+-ATPase (V-ATPase) inhibition. *J Biol Chem* **288**, 1385-1396 (2013).
715. Ong, C.A., *et al.* Three-gene immunohistochemical panel adds to clinical staging algorithms to predict prognosis for patients with esophageal adenocarcinoma. *J Clin Oncol* **31**, 1576-1582 (2013).

716. Langer, R., *et al.* Prognostic significance of expression patterns of c-erbB-2, p53, p16INK4A, p27KIP1, cyclin D1 and epidermal growth factor receptor in oesophageal adenocarcinoma: a tissue microarray study. *J Clin Pathol* **59**, 631-634 (2006).
717. Chan, D.S., Twine, C.P. & Lewis, W.G. Systematic review and meta-analysis of the influence of HER2 expression and amplification in operable oesophageal cancer. *J Gastrointest Surg* **16**, 1821-1829 (2012).
718. Paterson, A.L., *et al.* A systematic approach to therapeutic target selection in oesophago-gastric cancer. *Gut* **62**, 1415-24 (2013).
719. Waddell, T., *et al.* Epirubicin, oxaliplatin, and capecitabine with or without panitumumab for patients with previously untreated advanced oesophagogastric cancer (REAL3): a randomised, open-label phase 3 trial. *Lancet Oncol* **14**, 481-489 (2013).
720. Karapetis, C.S., *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359**, 1757-1765 (2008).
721. Diaz, L.A., *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540 (2012).
722. Solit, D.B. & Rosen, N. Resistance to BRAF inhibition in melanomas. *N Engl J Med* **364**, 772-774 (2011).
723. Lee, M.J., *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780-794 (2012).
724. Kimura, H., *et al.* Prognostic significance of EpCAM expression in human esophageal cancer. *Int J Oncol* **30**, 171-179 (2007).
725. Martin, I.G., Cutts, S.G., Birbeck, K., Gray, S. & Quirke, P. Expression of the 17-1A antigen in gastric and gastro-oesophageal junction adenocarcinomas: a potential immunotherapeutic target? *J Clin Pathol* **52**, 701-704 (1999).
726. Gires, O. & Bauerle, P.A. EpCAM as a target in cancer therapy. *J Clin Oncol* **28**, e239-240; author reply e241-232 (2010).